



A University of Sussex PhD thesis

Available online via Sussex Research Online:

<http://sro.sussex.ac.uk/>

This thesis is protected by copyright which belongs to the author.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Please visit Sussex Research Online for more information and further details

The biological role and mechanism of
action of *rbpA* and *carD* in *Streptomyces*
coelicolor A3 (2) and *Mycobacterium*
tuberculosis

Thesis submitted for the degree of Doctor of Philosophy

Heena Jagatia

Biochemistry and Biomedicine

School of Life Sciences

University of Sussex

Falmer, UK

September 2018

Declaration

I hereby declare that this thesis has not been and will not be submitted in whole or in part to another University for the award of any other degree.

.....

Heena Jagatia

Abstract

RbpA is unique to and conserved across the Actinobacteria and binds to the group I sigma factors to form an RNA polymerase holoenzyme complex resulting in stabilisation of the transcription initiation bubble. RbpA is composed of an unstructured N-terminus, RbpA core domain (RCD), Basic linker (BL) and Sigma Interaction Domain (SID). The function of the RbpA-SID is the only well characterised domain. Therefore, this study focuses on elucidating the function of the RbpA BL. A recently published full length RbpA crystal structure has shown that the RbpA BL is positioned to interact with the extended -10 promoter DNA. Alanine substitutions of the RbpA BL did show a small colony phenotype with actinorhodin production in *S. coelicolor*. The equivalent RbpA BL residue substitutions with alanine in *M. tuberculosis* were constructed and tested in vivo. The RbpA BL mutant (3KRA) failed to culture on solid media, but no growth inhibitory phenotype was present in liquid media. Additionally, RNA-seq of the BL mutant in both organisms revealed major transcriptional changes. Together these studies, suggest that the RbpA BL does not contribute to the essential function of RbpA and therefore RbpA may fulfil other functions. To explore these alternative functions, we constructed mutations in HrdB, a principle sigma factor in *S. coelicolor*, which prevented RbpA binding to RNAP/ σ^{HrdB} holoenzyme. Chromatin immunoprecipitation techniques followed by high-throughput ChIP sequencing of the HrdB mutant revealed that HrdB was capable of co-localising at all RbpA dependant promoters in the absence of RbpA binding.

CarD is another essential transcription factor in *Mycobacterium* spp which binds to RNAP/ σ^A holoenzyme. Several structural models have suggested that RbpA and CarD bind to the extended -10 promoter DNA from opposite sides of the double helix DNA. Therefore, to explore a potential relationship between RbpA and CarD, we constructed a degradation tag system to enable co-depletion of RbpA and CarD. This provides a genetic tool for studying essential genes without the use of antibiotic dependant inducible systems.

Acknowledgments

I would like to thank the University of Sussex for the school-funded studentship that I received for the PhD and BSMS for taking me on as an honorary member. I would like to thank my main supervisor Dr Mark Paget, for all of his advice, support and guidance throughout my PhD. I would like to thank everyone in the Paget lab, past members Dr Laurence Humphrey and Dr Heather Macklyne for their support and present members Dr Sophie Nicod and Hayley Greenfield. Additionally, I'd like to thank Tracy Nissan for constantly telling me to stay calm.

I would like to extend my gratitude to my co-supervisor, Dr Simon Waddell, who has played a large part in my successes through my PhD, there is no real way to put this into words. Thank you for believing in me. I would also like to thank all members of the Waddell lab, Dr Daire Cantillon for all the ridiculous drinking, Dr Leticia Muraro-Wildner for all the support during the final *M. tuberculosis* experiments, Dr Filomena Perrone for always being smiley. I'd also like to thank Dr Ben Towler for being my RNA-seq guru.

I could not have made the decision to study for a PhD without the amazing past academic supervisors, therefore I'd like to say thank you to Professor Philip Butcher, Dr Marcus Pond, Dr Nick Dorrell, Dr Abdi Elmi and Dr Neveda Naz.

Additionally, I'd like to thank all the friends I have made at the University of Sussex and Brighton and Sussex Medical School. I'd also like to acknowledge all the support staff (Sue Searle, Dr Amber Dorey, Dr Steve Pearce, Photis Charalambou and Juliet Kneller) who have worked endlessly, thank you to everyone who would autoclave 30 flasks and double autoclave MS agar at 4pm on a Friday because I needed it before midnight.

Outside of the scientific bubble I would like to thank my family (Gita, Pravin, Dharmendra, Pratiksha and Hitesh), friends and inspirational people I have met at Abbey College Birmingham, St George's University of London and the London School of Hygiene and Tropical Medicine. I would like to extend a thanks to Andrew Todd for always believing in me and putting up with me during the trials and tribulations of this PhD.

Contents

Abstract	iii
Acknowledgments	iv
Contents	v
Table of Figures	xi
Table of Tables	xiv
Abbreviations	xvi
1.1 Overview	2
1.1.1 Phylum Actinobacteria	2
1.1.2 Streptomyces genus	3
1.1.2.1 <i>Streptomyces coelicolor</i> A3 (2)	4
1.1.2.2 Genome of <i>S. coelicolor</i>	4
1.1.2.3 Life cycle of <i>S. coelicolor</i>	5
1.1.2.4 Regulation of antibiotic production in <i>S. coelicolor</i>	8
1.1.3 The <i>Mycobacterium</i> genus	11
1.1.3.1 <i>Mycobacterium tuberculosis</i>	12
1.1.3.2 Genome of <i>M. tuberculosis</i>	13
1.1.3.3 Pathogenesis of <i>M. tuberculosis</i> infection	14
1.1.3.4 Therapeutics	14
1.1.4 Overview of bacterial transcription	15
1.1.4.1 Bacterial DNA-dependent RNA polymerase	15
1.1.5 Structure and function of primary sigma factors	17
1.1.5.1 The $\sigma_{1.1}$ region	18
1.1.5.2 The σ_2 domain	19
1.1.5.3 The σ_3 domain	20
1.1.5.4 The σ_4 domain	21
1.1.5.5 Structure and function of anti-sigma factors	21
1.1.5.6 <i>Streptomyces coelicolor</i> sigma factors	22
1.1.5.7 <i>Mycobacterium tuberculosis</i> sigma factors	22
1.1.6 Transcription initiation	23
1.1.6.1 Transcription elongation	25
1.1.6.2 Transcription termination	26
1.1.7 RNAP binding regulatory factors	27
1.1.7.1 ppGpp	27
1.1.7.2 Crl	28

1.1.7.3	DksA.....	28
1.1.7.4	RapA (HepA)	29
1.1.7.5	RNA polymerase binding protein, RbpA	29
1.1.7.6	RNAP binding protein, CarD	34
1.1.7.7	Relationship between RbpA and CarD.....	37
1.1.8	Project aims.....	39
Chapter 2: Materials and Methods		40
2.1.1	Antibodies	43
2.1.2	Buffers and solutions.....	43
2.1.3	Strains.....	47
2.1.4	Plasmids.....	49
2.1.5	Antibiotics.....	50
2.1.6	Oligos.....	51
2.2	Growth selection and storage of bacterial strains.....	56
2.2.1	Liquid and solid media.....	56
2.2.2	Growth and storage.....	58
2.3	DNA manipulation	59
2.3.1	PCR	59
2.3.2	Gibson Assembly	61
2.3.3	Restriction digest.....	62
2.3.4	Gel electrophoresis	62
2.3.5	Gel purification	62
2.3.6	DNA dephosphorylation	63
2.3.7	DNA ligation.....	63
2.3.9.1	Small scale plasmid isolation from <i>E. coli</i>	63
2.3.9.2	Large Scale plasmid isolation from <i>E. coli</i>	63
2.3.9.3	Transformation of chemically competent <i>E. coli</i> DH5 α	64
2.3.9.4	Preparation of chemically competent <i>E. coli</i> (CaCl ₂ method)	64
2.3.9.5	Intergeneric conjugation of DNA into <i>S. coelicolor</i>	64
2.3.9.6	Electroporation of <i>M. tuberculosis</i>	65
2.4	Setting up cultures	65
2.5	Most probable number assay (MPN).....	66
2.6	Colony forming units (CFU) for <i>M. tuberculosis</i>	67
2.7	Protein purification	67
2.8.1	Gel filtration	68
2.8.2	Protein sample analyses via SDS-PAGE.....	68

2.8.3	Western blot analysis	68
2.9	RNA isolation from <i>S. coelicolor</i>	69
2.9.1	RNA isolation from <i>M. tuberculosis</i>	70
2.10	Chromatin immunoprecipitation.....	71
2.10.1	qPCR	72
2.11	S1 nuclease protection assay	73
2.11.1	Preparing the gel	73
2.11.2	S1 nuclease protection assay	73
Results chapter I: Investigating the role of the RbpA basic linker in <i>Streptomyces coelicolor</i> A3 (2)		75
3.1.0	Overview	76
3.1.1	Construction of the basic-linker mutants in <i>S. coelicolor</i>	77
3.1.2	Phenotype of RbpA BL mutants in <i>S. coelicolor</i>	78
3.1.3	Growth curves for the BL mutants in <i>S. coelicolor</i>	83
3.1.4	Expression levels of the BL mutants in <i>S. coelicolor</i>	85
3.1.5	Susceptibility of the RbpA BL mutants to rifampicin	87
3.1.6.1	RNA extractions in <i>S. coelicolor</i>	89
3.1.6.2	Sample preparation and the library preparation at Oxford genomics centre	90
3.1.7.1	RNA-sequencing bioinformatic analysis	93
3.1.7.2	FastQC and trimming of the reads.....	93
3.1.7.3	Mapping to the <i>Streptomyces coelicolor</i> A3 (2) genome.....	97
3.1.7.4	Assembling the transcriptome using Cufflinks	99
3.1.8	Analysis of the differential gene expression between RbpA ^{WT} and RbpA ^{3KRA}	101
3.1.8.1	Growth related genes are down-regulated in S401 pRT802:: <i>rbpA</i> ^{3KRA}	104
3.1	Discussion.....	112
3.2.1	The RbpA BL does not have an essential function.....	112
3.2.2	Increased susceptibility of RbpA BL mutants to rifampicin treatment	113
3.2.3	The RbpA BL mutant and Δ <i>rbpA</i> cause down-regulation of growth related genes	114
3.2.4	Future work	115
Results chapter II: Investigating the role of the RbpA basic linker in <i>Mycobacterium tuberculosis</i> .		116
4.0	Overview of the RbpA basic linker in <i>M. tuberculosis</i>	117
4.1.0	Construction of RbpA BL mutants in <i>M. tuberculosis</i>	117
4.1.1	Growth curves of the BL mutants in <i>M. tuberculosis</i>	120
4.1.2	Confirmation of complementation of the BL strains in <i>M. tuberculosis</i>	121
4.1.3	Characterising the growth phenotype of M233 (pMV306:: <i>rbpA</i> ^{3KRA})	123
4.2.0	Profiling the impact of <i>rbpA</i> ^{3KRA} substitutions on global gene expression patterns.....	126
4.2.1	Sample preparation and library preparation at Oxford genomics centre.....	127

4.2.2	FastQC and trimming of the reads.....	130
4.2.3	Mapping to the <i>M. tuberculosis</i> reference genome	131
4.2.4	Assembling the transcriptome using Cufflinks	132
4.2.5	Differential gene expression between M233 (pMV306:: <i>rbpA</i> ^{WT}) and M233 (pMV306:: <i>rbpA</i> ^{3KRA})	134
4.2.6	Minimal leaky expression from the repressed WT <i>rbpA</i> allele in M233 (pMV306:: <i>rbpA</i> ^{3KRA}) in the absence of PI	137
4.2.7	Biological insights from transcriptome profiling of the RbpA BL mutant	138
4.2.7.1	Hypergeometric analysis of genes induced in M233 (pMV306:: <i>rbpA</i> ^{3KRA}).....	138
4.2.7.2	Hypergeometric analysis of genes repressed in M233 (pMV306:: <i>rbpA</i> ^{3KRA}).....	140
4.2.7.2.1	Down-regulation of the DevR regulon by M233 (pMV306:: <i>rbpA</i> ^{3KRA}).....	141
4.2.7.2.2	RbpA ^{3KRA} down-regulates the DosR target, class II ribonucleotide reductase	147
4.2.8	The RNAP subunits are up-regulated in M233 (pMV306:: <i>rbpA</i> ^{3KRA})	148
4.3	Identification of potential transcription factors involved in response to BL mutation in <i>M. tuberculosis</i>	152
4.3.1	Analysis of the overlapping genes induced in M233 (pMV306:: <i>rbpA</i> ^{3KRA}) to transcription factor over expression datasets	152
4.3.2	The overlap of the down-regulated genes in M233 (pMV306:: <i>rbpA</i> ^{3KRA}) with the TFOE dataset	155
4.4	Discussion.....	156
4.4.1	RbpA ^{3KRA} displays an attenuated growth phenotype on solid agar.....	156
4.4.2	Impact of RbpA ^{3KRA} mutation on <i>M. tuberculosis</i> transcriptome	157
4.4.3	Future work	159
	Results chapter III: Investigating the role of RbpA in <i>S. coelicolor</i>	160
5.1.0	Overview	161
5.1.1	The importance of RbpA M85 residue interactions with σ^{HrdB}	161
5.1.1.1	Construction of the RbpA C-terminal M85 mutants.....	162
5.1.1.2	Phenotype of the RbpA ^{M85} mutants	163
5.2	Analysis of σ^{HrdB} mutants that cannot associate with RbpA	165
5.2.1	Overview	165
5.2.2.1	Construction of the σ^{HrdB} mutant for ChIP-analysis.....	166
5.2.2.2	The HrdB mutant association with RNAP using western blot analysis.....	167
5.2.2.3	Chromatin immunoprecipitation of HrdB ^{WT} and HrdB ^{4XR}	169
5.2.2.3.1	RbpA is not required for σ^{HrdB} interaction with SCO5281 promoter.....	169
5.2.2.4	Preparation of the ChIP-seq library	170
5.2.2.4.1	Bioinformatic analysis of ChIP-seq data	172
5.4	The genes that require RbpA in <i>S. coelicolor</i>	185

5.4.1	Analysis of the differential gene expression between S401 (pRT802) and S401 (pRT802:: <i>rbpA</i> ^{WT}).....	185
5.4.2	Analysis of genes up-regulated in S401 (pRT802).....	187
5.4.2.1	Genes induced by cell wall antibiotics are induced in S401 (pRT802).....	187
5.4.2.2	Increased levels of SigE associate with RNAP in Δ <i>rbpA</i>	189
5.4.3	Upregulation of genes in S401 (pRT802:: <i>rbpA</i> ^{3KRA}) and S401 (pRT802).....	192
5.4.3.1	Carotenoid biosynthesis operons, <i>crtEIVB</i> , <i>crtYTU</i> , <i>litRQ</i> and <i>litSAB</i> are up-regulated in S401 (pRT802) and S401 (pRT802:: <i>rbpA</i> ^{3KRA})	192
5.4.4	Down-regulation of genes in S401 (pRT802)	193
5.5	Discussion.....	193
5.5.1	The SID RbpA ^{M85} displays a clear phenotype	193
5.5.2	σ^{HrdB} unable to bind to RbpA is still able to associate with RNAP and promoters	194
5.5.3	Identification of targets dependent on RbpA	195
5.5.3.1	Up-regulation of genes in the absence of <i>rbpA</i>	195
5.5.4	Future work	196
	Results chapter IV: Investigating the relationship between <i>S. coelicolor</i> RbpA and CarD <i>in vivo</i>	198
6.1.0	Overview	199
6.1.1.0	Generation of stable RbpA-DAS and CarD-DAS strains.....	199
6.1.2.0	Validation of <i>sspB</i> -dependent depletion	204
6.1.2.1	Construction of the FLAG-DAS+4- tag strains	204
6.1.1.1	Depletion of RbpA-DAS+4 and CarD-DAS+4 in a <i>sspB</i> -dependant manner	206
6.1.2.2	The growth phenotype of the DAS+4 depletion strains are further amplified with Δ <i>rbpB</i> 208	
6.1.2.3	Depletion of the DAS tagged proteins in liquid culture	210
6.2.0	Testing for the relationship between <i>rbpA</i> and <i>carD</i>	213
6.2.1	Construction of the RbpA BL and CarD ^{W85A} integrative plasmids	213
6.3.0	Discussion.....	216
6.3.1	Induced degradation of RbpA ^{DAS+4} and CarD ^{DAS+4}	216
6.3.2	Validation of the DAS system using integrative RbpA BL mutants.....	217
6.3.3	Investigating a possible relationship between the DNA binding regions in RbpA and CarD 218	
	General discussion.....	219
7.0	Overview	220
7.1	Highlights of this study	220
7.2	Future directions	222
7.2.1	Understanding promoter structure which governs transcription initiation by RbpA and CarD 222	

7.2.2	Exploiting cryptic gene cluster for antibiotic production.....	223
	Appendices.....	225
	Appendix 1	226
	Appendix 2	227
	Appendix 3	263
	Investigating the dependency of RbpA and CarD on promoters	263
	Appendix 4	267
	References.....	271

Table of Figures

Figure 1.1 The life cycle of <i>Streptomyces coelicolor</i>	6
Figure 1.2 The structure of the four main compounds and Coelimycin P1 produced by <i>S. coelicolor</i>	11
Figure 1.2 Crystal structure of <i>Mycobacterium smegmatis</i> transcription initiation complex. .	16
Figure 1.3 The domain organisation for the four sigma factor groups.	17
Figure 1.4. Sliced view of the σ_2 domain binding to promoter DNA in <i>Thermus aquaticus</i>	20
Figure 1.5 The steps of transcription initiation.	24
Figure 1.6. Multiple sequence alignment of selected RbpA homologues.	30
Figure 1.7. The cryo-EM structure of <i>M. tuberculosis</i> σ^A /RbpA/up-stream fork DNA.....	32
Figure 1.8 Multiple sequence alignment of CarD homologues.	34
Figure 1.9. Structural model of the RPo/CarD complex in <i>T. thermophilus</i>	36
Figure 1.10. X-ray crystal structure of RbpA-transcription initiation complex with superimposed CarD.....	38
Figure 3.1. Cryo-EM structure of RNAP- σ^A holoenzyme with RbpA and complexed with upstream fork DNA in <i>M. tuberculosis</i>	77
Figure 3.2. Engineering of the RbpA BL mutants in S401.	80
Figure 3.3. Phenotype of the RbpA BL mutants on MS agar.	81
Figure 3.4. The phenotype of RbpA BL mutants on SMMS agar.	82
Figure 3.5. Growth curves of the BL mutants in <i>S. coelicolor</i>	83
Figure 3.6. Actinorhodin production in the BL mutants.	85
Figure 3.7. Expression levels of the BL mutant alleles in <i>S. coelicolor</i>	86
Figure 3.8. Sensitivity of the BL mutants in <i>S. coelicolor</i> to rifampicin.	88
Figure 3.9. Bioanalyzer results of the BL RNA samples from <i>S. coelicolor</i>	90
Figure 3.10. Overview of the library preparation carried out by the Oxford genomics centre.	92
Figure 3.11. FastQC results of sample S401 (pRT802:: <i>rbpA</i> ^{WT}) (replicate-1, paired read 2) before and after trimming.	96
Figure 3.12. Overview of the RNA-sequencing pipeline.	100
Figure 3.13. Quality of the RNA-seq comparing S401 (pRT802:: <i>rbpA</i> ^{WT}) and S401 (pRT802:: <i>rbpA</i> ^{3KRA}).	102
Figure 3.14. Heatmap generated using FPKM values for S401 (pRT802:: <i>rbpA</i> ^{WT}) and S401 (pRT802:: <i>rbpA</i> ^{3KRA}).	103
Figure 3.15. Down-regulation of ribosomal protein genes in S401 (pRT802:: <i>rbpA</i> ^{3KRA}) and S401 (pRT802).	105

Figure 3.16. Growth-related genes are down-regulated in S401 (pRT802:: <i>rbpA</i> ^{3KRA}) and S401 (pRT802).	106
Figure 3.17. The down-regulation of RNAP subunits in S401 (pRT802:: <i>rbpA</i> ^{3KRA}).....	107
Figure 3.18. Alignment of the 22 ribosomal protein gene promoters downregulated in S401 (pRT802:: <i>rbpA</i> ^{3KRA}) and S401 (pRT802).....	108
Figure 3.19. The down-regulation of the TCA cycle genes in S401 (pRT802:: <i>rbpA</i> ^{3KRA}) and S401 (pRT802).	109
Figure 3.20. Down-regulation of class Ia RNR, <i>nrdABX</i> operon.	110
Figure 3.21. Vitamin B12 dependant ribonucleotide reductase, <i>nrdJ</i> is down-regulated in S401 (pRT802:: <i>rbpA</i> ^{3KRA}) and S401 (pRT802).....	111
Figure 4.2. Growth of the <i>rbpA</i> BL mutants alongside <i>rbpA</i> ^{WT} and empty vector in <i>M. tuberculosis</i> in the absence of PI, measured by optical density.....	121
Figure 4.3. Growth of <i>M. tuberculosis</i> BL strains on solid media.	123
Figure 4.4. Growth rate assays for the RbpA BL mutants.	125
Figure 4.5. Bioanalyser results of the RbpA ^{WT} and RbpA ^{3KRA} RNA extractions from <i>M. tuberculosis</i>	126
Figure 4.6. Agilent 4200 Tapestation quality control results for the <i>M. tuberculosis</i> BL RNA samples performed by the Oxford Genomics Centre.	129
Figure 4.7. FastQC results of M233 (pMV306:: <i>rbpA</i> ^{WT}) replicate 1 before and after trimming.	131
Figure. 4.8. Pipeline for the processing of the RNA-sequencing samples in <i>M. tuberculosis</i>	133
Figure 4.9. Comparison of the biological replicates FPKM for <i>M. tuberculosis</i>	135
Figure 4.10. Differentially expressed genes in RbpA ^{3KRA} compared to RbpA ^{WT} in <i>M. tuberculosis</i>	136
Figure 4.11. Quantified RbpA ^{WT} and RbpA ^{3KRA} reads from the M233 (pMV306:: <i>rbpA</i> ^{3KRA}) strain.	137
Figure 4.12. The <i>devS/devR/Rv3134c</i> operon visualisation.	143
Figure 4.13. <i>devS/devR</i> FPKM values of the RbpA truncations, in <i>M. smegmatis</i>	144
Figure 4.14. The DevR regulon genes down-regulated in M233 (pMV306:: <i>rbpA</i> ^{3KRA}) compared to M233 (pMV306:: <i>rbpA</i> ^{WT}).	146
Figure 4.15. Down-regulation of <i>nrdZ</i> in by RbpA ^{3KRA}	147
Figure 4.16. The up-regulation of RNAP subunits in M233 (pMV306:: <i>rbpA</i> ^{3KRA}).	148
Figure 4.18. The <i>rpsJ</i> ribosomal protein operon.	152
Figure 4.19. Significant up-regulation of SigI and SigJ in M233 (pMV306:: <i>rbpA</i> ^{3KRA}).	155
Figure 5.1. The cryo-EM structure of RbpA-σ ^A complex in <i>M. tuberculosis</i>	162
Figure 5.2. Phenotypes of the RbpA M85 mutants on MS agar	164

Figure 5.3. Amino acid sequence alignment of <i>S. coelicolor</i> Group I, II and the IV sigma factors.	166
Figure 5.4. Western blot analysis of HrdB association with RNAP upon diamide stress.	168
Figure 5.5. RbpA is not required for σ^{HrdB} interaction with the SCO5281 promoter.	170
Figure 5.6. Pipeline for the library preparation of ChIP-seq samples.	171
Figure 5.7. Quality control of FASTQ files.	173
Figure 5.8. bigWig Histogram file visualised using the Integrated Genome Browser.	175
Figure 5.9. Validation of ChIP-sequencing for this study.	177
Figure 5.10. The dependency of tRNA genes on RbpA using ChIP-seq.	178
Figure 5.11. tRNA gene promoter alignments.	179
Figure 5.12. Alignment of gene promoters with less HrdB ^{4xR} binding.	184
Figure 5.13. Quality of the replicate RNA-Seq samples in <i>S. coelicolor</i>	186
Figure. 5.14 Motif conservation of the SigE targets.	189
Figure 5.15. Increased association of SigE in the S129 $\Delta rbpA$	191
Figure 6.1. The Meganuclease system to generate marker-less gene replacements in <i>S. coelicolor</i>	202
Figure 6.3. <i>sspB</i> -mediated depletion of RbpA in <i>S. coelicolor</i> J1915 (pIJ6902:: <i>sspB</i> , pRT802:: <i>rbpA</i> -FLAG-DAS+4).	205
Figure 6.4. <i>sspB</i> -mediated depletion of CarD in <i>S. coelicolor</i> J1915 (pIJ6902:: <i>sspB</i> , pRT802:: <i>carD</i> -FLAG-DAS+4).	206
Figure 6.5. The effects of <i>sspB</i> -dependent depletion of <i>rbpA</i> and <i>carD</i> on growth.	207
Figure 6.6. The effect of deletion of <i>rbpB</i> from the DAS+4 tagged <i>rbpA</i> and <i>carD</i>	209
Figure 6.7. The effects of <i>sspB</i> -dependent depletion of <i>rbpA</i> and <i>carD</i> on growth in a $\Delta rbpB$ null background.	210
Figure 6.8. <i>sspB</i> -dependant depletion of RbpA and CarD in liquid media.	212
Figure 6.10. The engineered elements of the <i>rbpA</i> -DAS+4 and <i>carD</i> -DAS+4 strains with pRT802.	214
Figure 6.11. Phenotype of RbpA BL and CarD ^{W85A} mutant alleles in the <i>sspB</i> -dependent co-depletion of <i>rbpA/carD</i> strain.	215
Figure. A3 Promoters tested for dependency on RbpA and CarD.	264
Figure. A4 Organisation of the S416 RbpA ^{DAS+4} /CarD ^{DAS+4} pSRE3::p2-3 (pIJ6902:: <i>sspB</i>) strain.	265
Figure A5. S1 nuclease mapping of the S416 (pIJ6902:: <i>sspB</i> ; pSRE3::p2-3) \pm thiostrepton. ...	266

Table of Tables

Table 2.1 List of strains used throughout the study.....	49
Table 2.2 List of plasmids used in this study.....	50
Table 2.3 List of antibiotics and additives used in this study including the stock/working concentrations in solid and liquid media.....	51
Table 2.4 List of primers designed for the BL mutations in <i>S. coelicolor</i>	51
Table 2.5 List of primers designed for the BL mutations in <i>M. tuberculosis</i>	52
Table 2.6 List of multiplex sequencing Illumina RNA-seq adapter oligos	52
Table 2.7 Primer designed for S1 nuclease protection assay	52
Table 2.8 Primers designed for the in vivo study of the ribosomal promoters in the RbpA/CarD-DAS depletion system.	53
Table 2.9 Primers designed for σ^{HrdB} association with RNAP using qPCR.....	53
Table 2.10 Adapter using for ChIP-seq library preparation	54
Table 2.11 List of the primers designed for the DAS dependant depletion system.	55
Table 2.12 Primers designed for the FLAG-DAS strains.	55
Table 2.13 Primers designed for validation of the DAS system.....	55
Table 2.14 Primers designed for <i>S. coelicolor</i> recombinant RNAP expression.....	56
Table 3.1. The number of reads contaminated with adapters.	94
Table 3.2. The number of reads kept and discarded after trimming of the <i>S. coelicolor</i> RNA samples.....	95
Table 3.3. Overview of the mapping ability of the sample reads to the <i>S. coelicolor</i> genome.....	98
Table 4.1. Quality control results and quantification of the <i>M. tuberculosis</i> RNA samples ...	127
Table 4.2. The number of reads kept and discarded after trimming.	130
Table 4.3. Overview of the sample reads using HISAT2.....	132
Table. 4.4. Hypergeometric analysis of genes induced RNA-seq gene list comparing RbpA ^{WT} and RbpA ^{3KRA}	139
Table 4.5. Hypergeometric analysis of the genes repressed in RbpA ^{3KRA} compared to RbpA ^{WT}	141
Table 4.6. Significantly enriched TF reulons in those genes up-regulated in M233 (pMV306::rbpA ^{3KRA}) vs M233 (pMV306::rbpA ^{WT}).	154
Table 4.7. Significantly enriched TF regulons in those genes down-regulated M233 (pMV306::rbpA ^{3KRA}) vs M233 (pMV306::rbpA ^{WT})	156
Table 5.1. Quality control of the ChIP samples in-house.....	172
Table 5.2. Percentage of reads mapped to the <i>S. coelicolor</i> genome	174

Table. 5.3. List of enrichment sites identified as significantly lower for HrdB ^{4xR} compared to HrdB ^{WT} using ChIP-seq.	182
Table 5.4. Promoter alignments of the SigE targets up-regulated in S401 (pRT802).	189

Abbreviations

Å: angstrom

ACT: Actinorhodin

BACTH: bacterial two-hybrid

BLAST: basic local alignment search tool

bp: base pairs

BSA: bovine serum albumin

CDA: calcium-dependent antibiotic

CFU: colony-forming unit

ChIP: chromatin immunoprecipitation

CTD: C-terminal domain

dH₂O: distilled water

DNA: deoxyribonucleic acid

dsDNA: double stranded deoxynucleic acid

ECF: extracytoplasmic function

FPLC: fast protein liquid chromatography

GC: guanine-cytosine

h: hour

HTH: helix-turn-helix

IGB: Integrated Genome Browser

IGV: Integrated Genome Viewer

INH: isoniazid

IPTG: Isopropyl β-D-1-thiogalactopyranoside

kb: kilobase pairs

kDa: kilodalton

MCS: multi-cloning site

MDR: Multi drug resistant

min: minute

MM: methylenomycin

MS: mannitol soya flour

NCBI: National Centre for Biotechnology Information

nt: nucleotides

NTD: N-terminal domain

NTT: N-terminal tail

OD: optical density

PCR: polymerase chain reaction

qPCR: quantitative polymerase chain reaction

RED: undecylprodigiosins

RID: RNAP interaction domain

RNA: ribonucleic acid

RNAP: RNA polymerase

RPc: RNA polymerase closed complex

RPo: RNA polymerase open complex

RT: room temperature

sec: seconds

SDS-PAGE: Sodium dodecyl sulphate polyacrylamide gel electrophoresis

SID: sigma-interaction domain

spp.: species

ssDNA: single stranded deoxynucleic acid

WT: wild-type

WHO: World health organisation

XDR: extremely drug resistant

ZBD: zinc-binding dom

Chapter 1: Introduction

1.1 Overview

Transcription initiation is catalysed by RNA polymerase (RNAP) which consists of a core enzyme composed of five subunits ($\alpha\beta\beta'\omega$) and a dissociable σ -factor which provides promoter specificity for transcription of a gene. Much of our understanding of transcription initiation originates from studies on *Escherichia coli* RNAP. However, studies have shown that *Mycobacterium tuberculosis* and other Actinobacteria such as *Streptomyces coelicolor* differ from this paradigm in requiring two additional essential transcription factors- RbpA and CarD.

S. coelicolor, is a Gram-positive bacterium found in the soil environment and was the model organism for the *Streptomyces* spp due to the production of four pigmented antibiotics. *M. tuberculosis* shares the phylum Actinobacteria with *S. coelicolor* and is an acid-fast bacilli.

This study focuses on the general transcription factors RbpA and CarD in both *S. coelicolor* and *M. tuberculosis*. It utilises the non-pathogenic *S. coelicolor* to investigate the transcriptional factors and the influence of these transcriptional factors on growth in *M. tuberculosis*.

1.1.1 Phylum Actinobacteria

Actinomycetes are the largest bacterial phylum and is comprised of Gram-positive bacteria with a high GC DNA content (Ludwig et al., 2005), ranging from 53.5% in *Corynebacterium* spp to 72% in *Streptomyces* spp and *Frankia* spp. Actinobacteria are found in diverse settings such as soil, air, human gastrointestinal tract and the marine environment (Barka et al., 2016). Morphologies vary widely across the class and strains are well known to produce extracellular enzymes and secondary metabolites (Barka et al., 2016; Schrempf, 2001).

The class of Actinobacteria are the largest within the Actinomycete phylum and includes the industrially important *Streptomyces* genus which is responsible for producing two-thirds of all known naturally derived antibiotics (Bentley et al., 2002; Papagianni, 2012) and *M. tuberculosis* that infects a third of the world's population (Doroghazi & Metcalf, 2013). The Actinobacteria phylum also includes the *Nocardia* spp, *Corynebacterium* spp, *Gordonia* spp, *Rhodococcus* spp, *Bifidobacterium* spp, *Leifsonia* spp, *Gardnerella* spp, *Frankia* spp and *Thermobifida* spp genus' (Barka et al., 2016). *Nocardia* spp are usually the causative agents of opportunistic infections in immunocompromised patients, usually acquired through inhalation

or percutaneous inoculation from the soil (Abreu et al., 2015) and is closely related to *Rhodococcus* which together belong to the family *Nocardiaceae*. The *Corynebacterium* genus includes the industrially important *C. glutamicum*, which produces valuable amino acids such as glutamate, lysine and tryptophan (Papagianni, 2012) whilst *C. diphtheriae* and *C. ulcerans* cause infections in humans (Barka et al., 2016). *Frankia* spp are involved in nitrogen-fixing (Ventura et al., 2007) and *Gardnerella* spp are the causative agent of bacterial vaginosis. Lastly, *Bifidobacterium* spp are non-spore forming, non-filamentous and have probiotic properties (Barka et al., 2016).

1.1.2 Streptomyces genus

The soil dwelling *Streptomyces* genus belongs to the Streptomycetaceae family and has over 3,000 species within the genera (Ochi, 1995). Streptomycetaceae are thought to have originated around 382 million years ago when the first land plants began to emerge and were responsible for the solubilisation of cell walls and surface material of fungi, plants and insects and thus were involved in early composting (Chater & Chandra, 2006; McDonald & Curriea, 2017). Unlike most bacteria, *Streptomyces* have a complex lifestyle and therefore it was difficult to classify in the early days, initially being classified as a fungus (Hopwood, 1999). The confusion remains with the name *Streptomyces* which means ‘twisted or chain-like fungus’. This genus is well known for its clinically relevant production of antibiotics and anti-tumour compounds (Chater, 2016). *S. coelicolor* has been the model organism for genetic analysis of differentiation and antibiotic production in the *Streptomyces* genus since the early 1960’s. The developmental differentiation of *S. coelicolor* is defined into two classes by mutants blocking distinct stages of the life cycle (Bush et al., 2013). The *bld* mutants are unable to form aerial hyphae and have a characteristic shiny “bald” phenotype, whilst the wild-type colony appears “fuzzy” in appearance (Flärdh & Buttner, 2009). The *whi* mutants can form aerial hyphae but unable to produce mature chains of spores, therefore the colonies appear white on solid media compared to a grey pigment in a wild-type colony (Flärdh & Buttner, 2009). However, *S. coelicolor* is unable to sporulate in liquid media and therefore *Streptomyces venezulae* has recently been used as the model organism to complete the developmental cycle (Chater, 2016).

1.1.2.1 *Streptomyces coelicolor* A3 (2)

Streptomyces coelicolor A3 (2) is a filamentous soil bacterium and the most well studied organism as it best represents the *Streptomyces* genus (Hopwood, 1999). *S. coelicolor* can produce a rich array of secondary metabolites as a result of the biosynthetic gene clusters (Thakur et al., 2013). *S. coelicolor* produces at least four antibiotics, actinorhodin (ACT) and undecylprodigiosins (RED), calcium-dependent antibiotic (CDA) and methylenomycin (MM) (Liu et al., 2013; Fig. 1.2).

1.1.2.2 Genome of *S. coelicolor*

S. coelicolor has a large linear genome spanning 8,667,507 bp with a 72.12 % GC content which is characteristic of the *Actinobacteria* phylum (Bentley et al., 2002). The coding density of the genome centralises genes involved in cell division, DNA replication, transcription, translation and amino acid biosynthesis into the core and are located close to the *oriC* which is also centralised in the chromosome (Bentley et al., 2002). The density is slightly decreased in the distal regions where accessory genes such as those coding for specialised metabolites and hydrolytic exoenzymes are located (Bentley et al., 2002). The *S. coelicolor* genome has a predicted 7,825 genes, giving it high coding potential. In addition to this large genome, are two plasmids: 350 kb linear, low copy number SCP1 and the 31.4 kb circular, low copy number SCP2 (Kinashi & Shimaji-Murayama, 1991). An estimated 965 proteins are associated with regulation, including 65 sigma factors which bind to promoter sequences affecting promoter specific transcription initiation upon detection of specific stresses (Bentley et al., 2002). The genome sequence also revealed a large number of two-component systems, 85 sensor kinases, 79 response regulators and approximately 53 sensor-regulator pairs. Large numbers of transcriptional regulation proteins have also been identified and belong to some of the most widely known families such as TetR, GntR, IclR and MerR. Aside from transcriptional regulators, 614 proteins are involved in transport of substrates and metabolites, such as the ABC transporters, permeases and ATP binding proteins, and 819 secreted proteins which allow *S. coelicolor* to grow in such a complex environment (Bentley et al., 2002).

1.1.2.3 Life cycle of *S. coelicolor*

The life cycle of *S. coelicolor* and other *Streptomyces* is unlike most other bacteria, it resembles the life cycle of fungi. The dormant spores are activated upon detection of an appropriate nutrient source (Fig. 1.1A; Sigle et al., 2015), one or two extensions termed germ tubes extend from the spore during germination (Fig. 1.1B; Wolanski et al., 2011). Growth occurs at the hyphal tip ends (Flärdh, 2003a) and is unlike other rod-shaped bacteria such as *E. coli*, where cell elongation is orchestrated by the helical cytoskeleton formed by bacterial actin homologue MreB and new peptidoglycan material is added to the lateral wall and the cell poles remain inert after cell division (Flärdh & Buttner, 2009). Most *Actinobacteria* do not code for *mreB* genes and yet still grow by extension at the cell poles. However, *S. coelicolor* does contain a complete *mre* gene cluster, although tip extension in *S. coelicolor* does not depend on Mre proteins (Sigle et al., 2015; Flärdh & Buttner, 2009). *Actinobacteria* without *mreB* depend on a coiled coil protein called DivIVA. DivIVA is an essential protein in *S. coelicolor* and is implicated in tip extension, branching and cell shape (Fig. 1.1C). Partial depletion of DivIVA causes defective hyphal growth and without DivIVA spores are unable to produce germ tubes (Flärdh, 2003b; Flärdh & Buttner, 2009). The hyphae develop into vegetative mycelium which extend and anchor the colony into the solid agar (Fig. 1.1D).

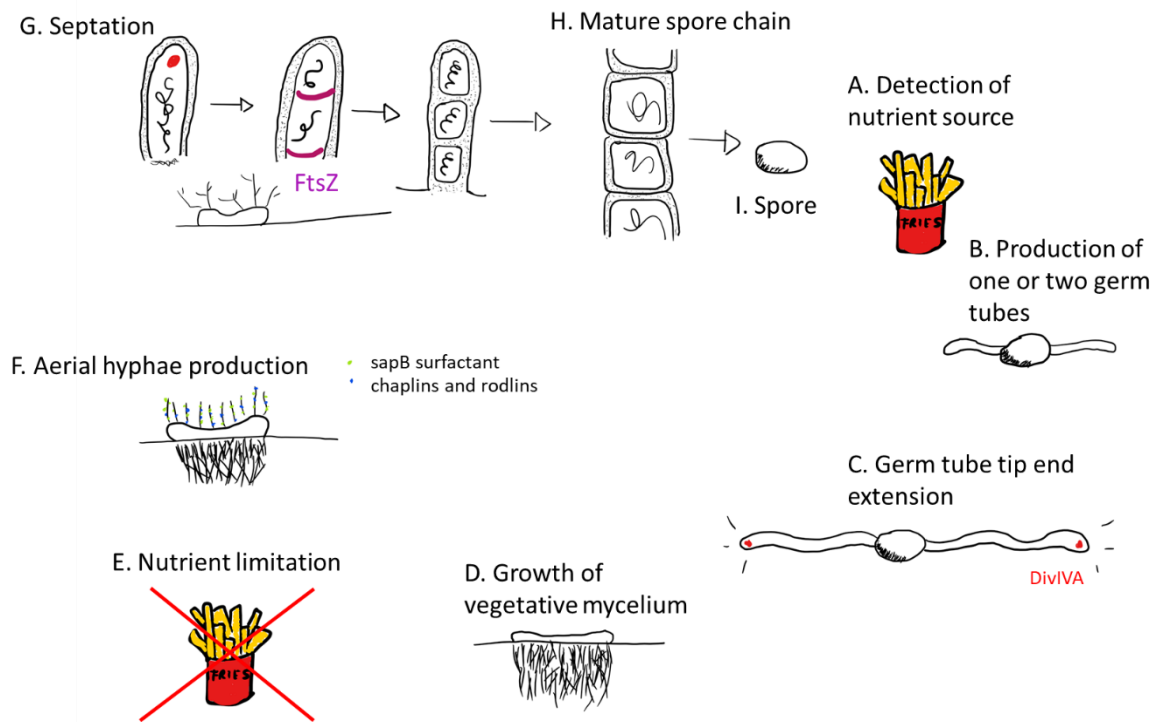


Figure 1.1 The life cycle of *Streptomyces coelicolor*. (A) Dry spore detects nutrient source, (B) one or two germ tubes extend from the germinating spore, (C) Germ tube tip extension regulated by DivIVA, (D) Formation of vegetative hyphae anchoring the colony into the solid agar, (E) Nutrient limitation detected, (F) Aerial hyphae formation, (G) Multi-genomic sporogenic cell without septation, then septation by FtsZ, and single chromosomal copies packed into each cell by ParAB and Ftsk (H) Spore maturation and increased cell wall thickening (Newton & Fahey, 2008) and (I) individual dry spores.

When the hyphae detect nutrient limitations (Fig. 1.1E), they begin to extend as aerial hyphae (Fig. 1.1F). The vegetative hyphae are coated in a hydrophobic sheath and a surfactant peptide SapB which reduces the surface tension allowing extension in the air and gives *S. coelicolor* its characteristic fuzzy appearance (Flärdh & Buttner, 2009). SapB is a Lantibiotic-like peptide with no antibiotic activity, derived from *ramS* encoded pre-peptide which undergoes extensive RamC-mediated post-translational modification (Kodani et al., 2004). SapB is not produced in minimal medium, which suggests that *S. coelicolor* may utilise a SapB-independent pathway. The SapB-independent pathway is mediated by the chaplin and rodlin proteins which constitute to the hydrophobic sheath (Flärdh & Buttner, 2009). Chaplin proteins are encoded

by the *chpABCDEH* operon and are an assortment of long and short chaplins which self-assemble into amyloid-like filaments at the water/air interface. ChpA-C encode large chaplins, with an amino-terminal Sec signal peptide, two chaplin domains separated by a linker and a C-terminal which targets the chaplins for covalent attachment to the cell wall of the aerial hyphae via a sortase enzyme (Claessen et al., 2003). ChpD-H encode short chaplins and consists of a chaplin domain followed by a Sec signal peptide (Claessen et al., 2003). The expression of *chp* genes is activated before aerial hyphae formation and strains that have defective SapB or chaplin proteins exhibit a *bld* phenotype (Capstick et al., 2007). *bld* mutants are unable to erect aerial hyphae and therefore appear 'bald' and lack the characteristic fuzzy phenotype (Flärdh & Buttner, 2009). Multiple *bld* genes are important for aerial hyphae production, for example, *bldA* encodes for the only tRNA that translates the rare codon TTA in *Streptomyces* spp. BldA produces mature tRNA during later stationary phase and the presence of the rare TTA codon in all four-antibiotic biosynthetic cluster, influences the production of antibiotics (McCormick & Flärdh, 2012). Another component of the hydrophobic sheath is the rodlin proteins, RdlA and RdlB. Strains lacking RdlA and RdlB still exhibit normal aerial hyphae, however the small basketwork filaments are disrupted (Claessen et al., 2003; Fig. 1.1F).

Next the aerial hyphae undergo a developmental change from aerial hyphae extension to septation. During the early aerial hyphal growth, hyphae are long, non-septated and an apical compartment known as the sporogenic cell forms. The sporogenic cell synthesises more than 50 copies of the chromosome, and arrests extension of the hyphae for cell division to take place (Claessen et al., 2003). The developmental stage of cell division is controlled by a bacterial tubulin homologue, FtsZ which initiates septation. FtsZ assembles into helical filaments which are modelled in Z-rings that direct septation and recruitment of other cell-division associated proteins. *S. coelicolor* *ftsZ* null mutants can produce aerial hyphae but are unable to produce spores (McCormick et al., 1994). The sporogenic cell can contain more than 50 copies initially which span the length of the cell, however the DNA then segregates into unigenomic spores followed complete partitioning into individual nucleoids (Fig. 1.1G), and this process is controlled by ParAB and FtsK. ParAB encodes for a cytoskeletal ATPase ParA and the DNA-binding protein ParB, which binds to *parS* sites. Disrupted *parAB* does not display any growth phenotype but *parB* affects chromosomal partitioning in aerial hyphae (Jakimowicz et al., 2006). ParB binds to the numerous *parS* sites located around the *oriC* which enables ParB

to assemble into a large nucleoprotein complex present at foci which are regularly distributed in a sporogenic cell ready for the formation of prespore compartments (Jakimowicz et al., 2006). FtsK is localised at septa and is important for pumping DNA through the septa, *ftsK* mutants demonstrate irregular DNA content and is a result of chromosomal DNA trapping at the septa causing chromosomal rearrangements and deletions (Claessen et al., 2003).

Lastly, the prespore matures by thickening of the wall and rounding into a spherical spore (Fig. 1.1G, 1.1H). As *S. coelicolor* contains the complete *mre* cluster, MreB regulates the assembly of the mature spore wall. *S. coelicolor* spore *mreB* null mutants exhibit thin walls and irregular shaping (Sigle et al., 2015). The process of sporulation requires many proteins, *S. coelicolor* mutants that fail to produce a grey pigment have been defined as demonstrating a 'whi' phenotype (Flärdh & Buttner, 2009).

1.1.2.4 Regulation of antibiotic production in *S. coelicolor*

The genome sequence of *S. coelicolor* revealed more than 18 gene clusters associated with secondary metabolite production located on the distal arms of the chromosome (Bentley et al., 2002). The secondary metabolites include: antibiotics, siderophores, terpenoids, lipids and other molecules (Bentley et al., 2002; Čihák et al., 2017). The structures of as many as 30% of the secondary metabolites are known, but the remaining 70% are produced by 'cryptic gene clusters', which are not activated by using common culturing conditions in the laboratory thus far. *S. coelicolor* has been known to produce four antibiotics which are discussed in further detail below.

Methylenomycin (MM) was the first to be discovered in *S. violaceoruber* in 1974 (Haneishi et al., 1974) and the first known antibiotic identified in *S. coelicolor*. The *mmy* gene cluster is located on the linear plasmid SCP1 and is regulated by small autoregulatory molecules, methylenomycin furans (MMFs) (Liu et al., 2013). The gene cluster is composed of three furan biosynthesis genes (*mmfL*, *mmfH* and *mmfP*) and two regulatory genes (*mmyR* and *mmfR*) (Liu et al., 2013). A proposed theory is that an increase in the MM concentration causes repression of the *mmyR* and the *mmfR* bidirectional promoter, which causes increased biosynthesis of the MMFs relieving the MmyR/MmfR-mediated repression of *mmyB* which then activates the MM biosynthetic genes (Liu et al., 2013).

The second antibiotic discovered in *S. coelicolor* was Actinorhodin (ACT) identified in the late 1960s and displayed inhibitory action against Gram-positive bacteria (Rudd and Hopwood, 1979), the mass spectra similarities of the isolated actinorhodin preparation from Brockmann (1966), Wright and Hopwood (1976) identified the secondary metabolite in *S. coelicolor*. ACT is very soluble and blue-pigmented at a pH above 7, whilst it appears red and highly insoluble below pH 7 (Rudd & Hopwood, 1979). The *act* gene cluster is activated by the transcription factor ActII-ORF4, and its promoter is regulated through multiple pathways (van Wezel & McDowall, 2011). For example, AbsA2 is a repressor of ActII-ORF4 and the *absA* operon consists of the sensor kinase, *absA1* which phosphorylates the response regulator, *absA2* and mutants of either of these causes antibiotic production (Sheeler et al., 2005). DasR is another repressor of actII-ORF4 belonging to the GntR transcriptional regulatory family, which regulates the GlcNAc regulon and GlcNAc causes growth arrest of *S. coelicolor* in rich media (Rigali et al., 2008). Glucosamine-6-phosphate inhibits DasR binding to DNA, *dasR* mutants display increased production of ACT (Rigali et al., 2008; Wiatek-Połatyńska et al., 2015). The soil contains very poor concentrations of nitrogen and carbon for *S. coelicolor* to survive. Therefore, the nitrogen and carbon source N-acetylglucosamine (GlcNAc) which forms bacterial peptidoglycan and chitin monomers is an appropriate energy source. The actinorhodin-associated transcriptional regulator, AtrA is an activator of actII-ORF4 (Uguru et al., 2005), AtrA may adjust acetyl-CoA metabolism which provides the precursors for polyketides and thus antibiotic production (van Wezel & McDowall, 2011). Disruption of *atrA* causes a reduction in the production of actinorhodin and the levels of mRNA are also reduced, suggestive that AtrA is a transcriptional regulator of actII-ORF4 (Uguru et al., 2005).

Red complex of undecylprodigiosin (RED), is a red pigmented antibiotic and was the third antibiotic to be identified in *S. coelicolor* (Rudd & Hopwood, 1980). Prodiginine production in *S. coelicolor* is controlled by the *redA-E* biosynthetic gene cluster. The *red* pathway is regulated by the response regulator, RedZ which activates expression of a direct activator of the *red* biosynthesis gene cluster (Liu et al., 2013).

Calcium-dependant antibiotic (CDA) demonstrates antimicrobial action in the presence of calcium and was the fourth antibiotic identified in *S. coelicolor* (Lakey et al., 1983). The two-component system, *absA1/absA2* is proposed to regulate *cdaR* which activates the *cda*

biosynthesis gene cluster (Liu et al., 2013). The phosphorylation of *absA2* causes repression of the *cdaR* promoter (Sheeler et al., 2005).

bldA mutants are viable but morphological differentiation and the production of all four, ACT, MM, CDA and RED secondary metabolites is inhibited (Chater & Chandra, 2008). The *bldA* tRNA recognises the TTA codon present in *actII-ORF4*, *mmyB*, *mmfL* and *redZ*. Therefore, *bldA* mutants are unable to produce ACT, RED and MM, whilst substitution to an alternative leucine codon restores ACT production (Leskiw et al., 1991).

Coelimycin P1 is a yellow-pigmented metabolic product of the 'cryptic *cpk* gene cluster' and was discovered in a *S. coelicolor* M1148 strain which lacks the *act*, *red* and *cda* antibiotic biosynthetic gene clusters. Mutations in the RNAP, *rpoB* gene prevents production of ACT, RED and CDA and thus reduces the competition for precursors, allowing expression of the cryptic *cpk* gene cluster (Gomez-Escribano et al., 2012).

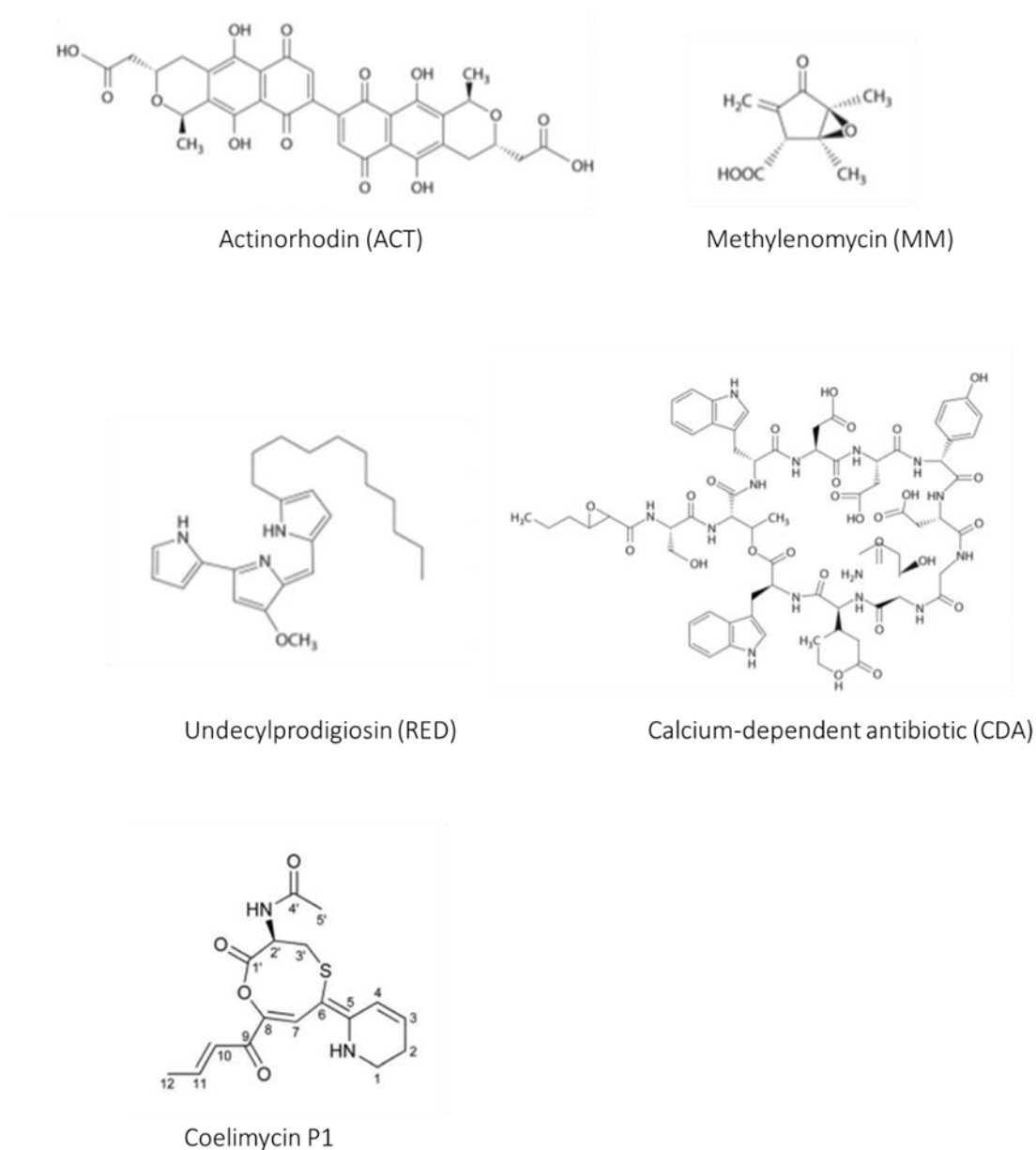


Figure 1.2 The structure of the four main compounds and Coelimycin P1 produced by *S. coelicolor*. Actinorhodin (ACT), methylenomycin (MM), undecylprodigiosin (RED), calcium-dependent antibiotic (CDA) are the four main compounds produced by *S. coelicolor* whilst coelimycin P1 is produced in the absence of the ACT, MM, RED and CDA gene clusters (Gomez-Escribano et al., 2012; G. Liu et al., 2013).

1.1.3 The *Mycobacterium* genus

There are around 170 *Mycobacterium spp* belonging to the *Mycobacterium* genus of the *Mycobacteriaceae* family (Forbes, 2017). Most of the mycobacterial genus are found in the

environment and have been isolated from soil, water, domestic/wild animals (Falkinham, 2009). The infectious *M. tuberculosis* complex includes *Mycobacterium tuberculosis*, *Mycobacterium bovis*, *Mycobacterium africanum*, *Mycobacterium microti* and *Mycobacterium leprae*, and the laboratory mutant strain of *M. bovis* known as *M. bovis* var BCG. BCG is also the vaccine strain that was first introduced in 1921 and has been used to vaccinate against childhood tuberculosis (Dye, 2013). *M. canetti* and *M. africanum* are closely related to *M. tuberculosis* and has been isolated from African patients. *M. bovis* affects humans and cattle, whilst *M. caprae* has been isolated from goats and *M. microti* has been isolated from rodents and immunocompromised patients (Brosch et al., 2002; Forrellad et al., 2013). *Mycobacteria* are classified as a Gram-positive bacterium as they lack an outer membrane, however *Mycobacteria spp* do share some resemblance to Gram-negative bacteria by the failure to retain the Gram stain, and contain porins in the outer membrane (Hett & Rubin, 2008). *Mycobacteria spp* are referred to as acid-fast because of the carbol fuchsin dye retention (Hett & Rubin, 2008). The presence of a lipid-rich outer membrane is the major determinant of *Mycobacteria* ecology and epidemiology as it contributes to the cells hydrophobicity, impermeability (Falkinham, 2009). The impermeability of the cell wall renders *Mycobacteria spp* naturally resistant to many antimicrobial drugs (Hett & Rubin, 2008).

1.1.3.1 *Mycobacterium tuberculosis*

M. tuberculosis was first identified by Robert Koch, a German physician and microbiologist in 1882 and who was subsequently awarded the Physiology or Medicine Nobel prize in 1905 for his discovery (nobelprize.org). *M. tuberculosis* are curved non-motile, acid-fast rods and obligate aerobes that are naturally resistant to acid, alkalis and dehydration with a doubling time of 22 h (McMurray, 1996). *M. tuberculosis* is the causative agent of tuberculosis infection that caused 9 million new cases and 1.7 million deaths in 2016 (WHO, 2017). *M. tuberculosis* was classified as one of the top ten causes of deaths worldwide in 2016 (WHO, 2017) and infects 1 in 3 people worldwide, of these people 3-5% develop symptomatic tuberculosis disease (Smith, 2003). Tuberculosis is not an eradicated disease in the UK despite public misconceptions, there are still ~6,000 new cases annually in the UK (Public Health England, 2017). The complex bacterial cell wall composed of arabinogalactan and mycolic acids has proven to be difficult to penetrate therapeutically which goes in some way to explaining

continued difficulty in eradicating tuberculosis. Some of the risk factors for tuberculosis include, homelessness, HIV, immigration, drug use and close living arrangements (Narasimhan et al., 2013). *M. tuberculosis* is spread through air-borne droplets, usually coughed into the air and transmitted person-to-person. Symptoms of tuberculosis include: coughing blood, severe tiredness, night sweats and fever. The most common site of infection is the lungs (pulmonary disease), however infections may also be seen in the bones/brain/lymph nodes/intestines and other extra-pulmonary sites (Pollett et al., 2016).

1.1.3.2 Genome of *M. tuberculosis*

The reference *M. tuberculosis* H37Rv strain was originally isolated from a pulmonary tuberculosis patient in 1905. The *M. tuberculosis* genome was sequenced in 1998, and at the time was the largest genome sequenced after *E. coli* (Cole et al., 1998). The genome is 4,411,529 bp with 65.6% GC content and 3,924 genes (Cole et al., 1998). The start codon of *dnaA* marks the numbering of genes (Cole et al., 1998). A single ribosomal RNA operon is located 1.5 kb from the *oriC*, which is unusual as most bacteria have more than one rRNA operon. The lack of multiple *rrn* operons may contribute to the slow growth rate of *M. tuberculosis* (Cole et al., 1998). The *M. tuberculosis* genome encodes more than 214 regulatory proteins (Minch et al., 2015; Rustad et al., 2014; Turkarslan et al., 2015), including 13 sigma factors, 11 two-component systems and a large family of eukaryotic-like serine/threonine protein kinases involved in signal transduction. The genome sequence includes 14 potential drug-efflux systems and ABC transporters that may enable the bacilli to be naturally resistant to many antibiotics (Cole et al., 1998). There are 250 distinct enzymes involved in fatty acid metabolism compared to 50 in *E. coli*. Lastly, 10% of the genome encodes members of the PE and PPE protein families. The names PE and PPE derive from Pro-Glu and Pro-Pro-Glu repeats, respectively. The *M. tuberculosis* H37Rv, reference strain codes for 99 PE genes that are scattered throughout the genome and that share a highly conserved N-terminal domain of 90-100 amino acids in length (Brennan, 2017). PPE proteins also share highly conserved N-terminal domains of 180 amino acids in length and PE and PPE genes are often co-transcribed. The functions of PE and PPE families remain unclear; however, a large number of the PE/PPE proteins display immuno-modulatory properties (Brennan, 2017).

1.1.3.3 Pathogenesis of *M. tuberculosis* infection

M. tuberculosis is aerosolised when coughed out of the lungs of an infected individual, inhalation of the infectious droplets allows entry of the bacilli into the alveolar of the new host. Alveolar macrophages phagocytose the bacilli, the host macrophages can clear the infection. Alternatively, the bacilli can reside within the macrophage vacuoles, avoiding phago-lysosomal fusion, and begin to divide (Bermudez & Goodman, 1996). Bacilli antigen detection by dendritic cells causes an immune response to be mounted. There are many mechanisms by which *M. tuberculosis* evades clearance by the host which have not been discussed here. Infections result in the formation of an aggregate of immune cells known as a granuloma, the hallmark of mycobacterial infection. Bacilli are able to survive within these granulomas and persist for decades, the dysregulation of the immune system can then lead to the granuloma progression and disease (Guirado & Schlesinger, 2013).

1.1.3.4 Therapeutics

The standard tuberculosis treatment is a six-month long regimen consisting of four antibiotics (WHO, 2015): isoniazid, rifampicin, pyrazinamide and ethambutol for the first two months and then isoniazid and rifampicin for the last four months (Vilchèz et al., 2013). This is a long drug regimen, which can lead to poor patient compliance or incomplete treatment, all factors associated with the rise of antimicrobial resistance. The symptoms of tuberculosis are non-specific and therefore it can take a long time for diagnosis. Isoniazid (INH) has been used clinically since 1952 for the treatment of tuberculosis. Isoniazid is a prodrug that enters by passive diffusion into *M. tuberculosis* (Bardou et al., 1998), this is activated by catalase peroxidase, *katG* in *M. tuberculosis*. The INH-NAD adduct produces radicals that can react with the mycobacterial cell and inhibit mycolic acid synthesis resulting in cell death of replicating *M. tuberculosis* (Winder & Collins, 1970). Pyrazinamide (PZA) was recognised as an anti-tuberculosis drug in 1952. Pyrazinamide is converted to pyrazinoic acid by nicotinamidase encoded by *pncA* and is only active against bacilli in an acidic pH or within macrophages (Scorpio & Zhang, 1996). Ethambutol was introduced in 1961, it inhibits the cell wall arabinogalactan biosynthesis (Deng et al., 1995). Rifampicin is a potent anti-tuberculosis drug that was introduced into the anti-tuberculosis regimen in 1968. Rifampicin inhibits RNA

synthesis by binding to the β subunit of DNA-dependent RNA polymerase whilst resistance is associated with mutations in *rpoB* (Campbell et al., 2001; Hartmann et al., 1967).

1.1.4 Overview of bacterial transcription

Transcription is the process by which information from DNA is transferred to RNA. The first phase of transcription is initiation and begins with DNA-dependent RNA polymerase in bacteria binding to a promoter DNA sequence found upstream of a gene. Transcription initiation is tightly controlled by positive and negative regulators of RNAP (Murakami & Darst, 2003). The second phase is elongation which requires disassociation of the σ subunit, and the elongation complex progresses along the DNA template whilst producing nascent RNA as well as regular pausing to regulate the accuracy of DNA transcription. The final phase is termination, RNAP dissociates from the DNA template and releases the nascent RNA.

1.1.4.1 Bacterial DNA-dependent RNA polymerase

Core RNA polymerase consists of a catalytic core composed of five subunits, two alpha subunits (α), one beta subunit (β), one beta prime subunit (β') and one omega subunit (ω ; Fig. 1.3). The molecular mass of core RNA polymerase is 400 kDa and is conserved across all cellular organisms making it a well-studied structure (Archambault & Friesen, 1993). Electron microscopy was used to determine the first three-dimensional structure of RNA polymerase in *E. coli* (Darst et al., 1989). This was followed by the X-ray crystallography of the three-dimensional *Thermus aquaticus* RNA polymerase structure, resolution of 3.3 Å (Zhang et al., 1999). The RNA polymerase structure is 150 Å long, 115 Å tall, and 110 Å wide (Zhang et al., 1999). The structure suggests a 'crab claw' like shape, one arm represents the β subunit, whilst the second arm is the β' subunit, with an internal channel running along the length of the 'claw' (Darst et al., 1989). The β and β' subunits make extensive interactions with one another, the active site carries Mg^{2+} which exists between the two subunits (Zhang et al., 1999). The α subunit is composed of two domains; an amino terminal domain (α NTD) linked by a flexible linker to a carboxyl-terminal domain (α CTD). The α NTD is the domain which dimerises whilst the α CTD domain only dimerises weakly (Ebright, 1995). The α NTD dimer aids in the assembly of β and β' and is not involved in catalysis (Zhang et al., 1999). The α CTD may contact the

upstream promoter DNA, and deletion of α CTD limits transcription at highly expressed ribosomal *rrnB* promoters (Ross et al., 1993). The flexible linker may allow the α CTD to contact the upstream promoter element of -35. The ω subunit encoded by *rpoZ*, interacts exclusively with the C-terminal tail of β' subunit, and may play a chaperone like function in the assembly of the final RNA polymerase structure (Mukherjee & Chatterji, 1997). The RNA polymerase inhibitory site for rifampicin binding is roughly 20 Å from the Mg^{2+} chelating active site (Zhang et al., 1999), -2 to -3 initiating nucleotides of the template DNA covalently attach to rifampicin (Mustaev et al., 1994). The structure of RNA polymerase consists of a sixth dissociable, σ -factor. The RNAP/ σ -factor complex is referred to as holoenzyme (Fig. 1.3). The sigma factor provides promoter specific transcription initiation of specific genes.

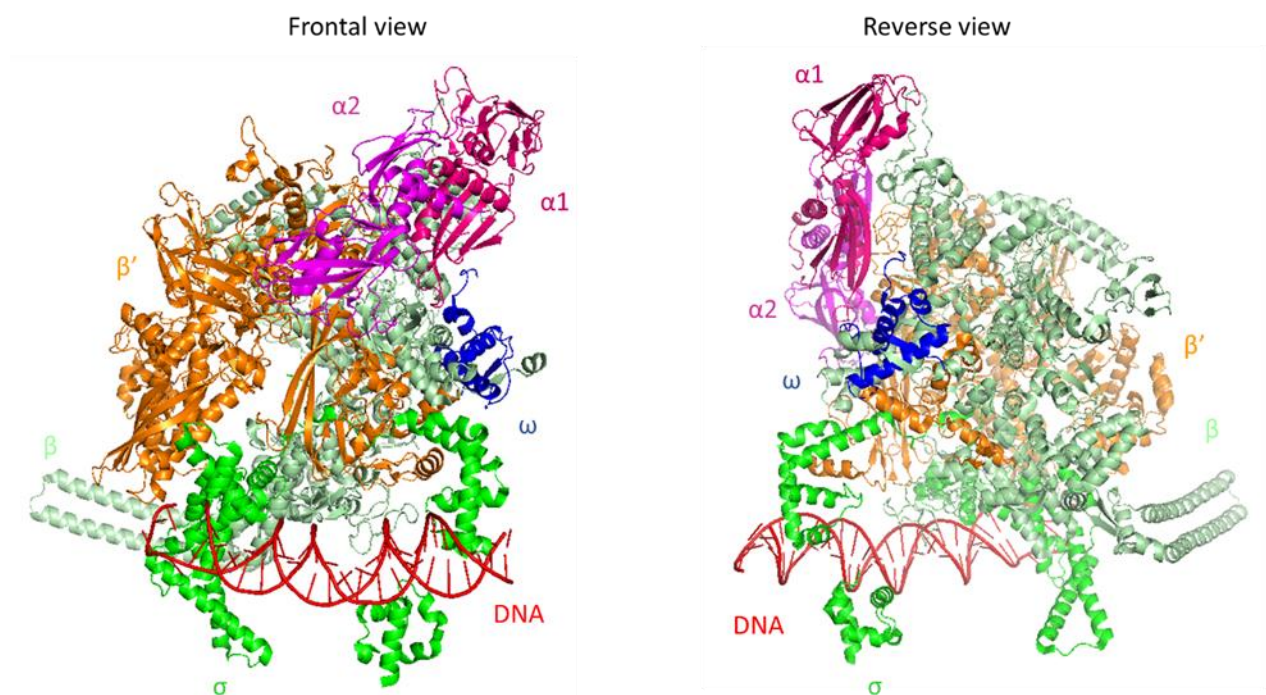


Figure 1.2 Crystal structure of *Mycobacterium smegmatis* transcription initiation complex. Structure of *M. smegmatis* RNA polymerase transcription initiation complex with σ^A . The β subunit in mint, β' in orange, $\alpha 1$ in pink, $\alpha 2$ in cyan, ω in blue and σ^A in green. The structure was downloaded from RCSB PDB, accession number 5TW1 and visualised with PyMOL (version 2.0.7; Hubin et al., 2017).

1.1.5 Structure and function of primary sigma factors

Sigma factors were first discovered by Burgess and Dunn, who described that σ factors enhanced RNA synthesis dependant on the DNA template (Burgess et al., 1969). Most bacterial σ factors that fall into one of two main families exemplified by the model of σ factors in *E. coli*: σ^{70} and σ^{54} . The σ^{54} family are phylogenetically distinct and respond to environmental stimuli (Paget & Helmann., 2003; Wigneshweraraj et al., 2008), and are not considered further here. The σ^{70} family of sigma factors can be divided into 4 distinct groups (Fig. 1.4). Domains σ_2 , σ_3 , and σ_4 interact with the -10 (TATAAT), extended -10 and -35 (TTGACA) respectively, and RNAP (Paget, 2015).

Group I are primary sigma factors which direct the transcription of essential genes associated with growth and have four domains, $\sigma_{1.1}$, σ_2 , σ_3 , σ_4 joined by flexible linkers. Most bacteria encode for at least one group I sigma factor (Murakami & Darst, 2003).

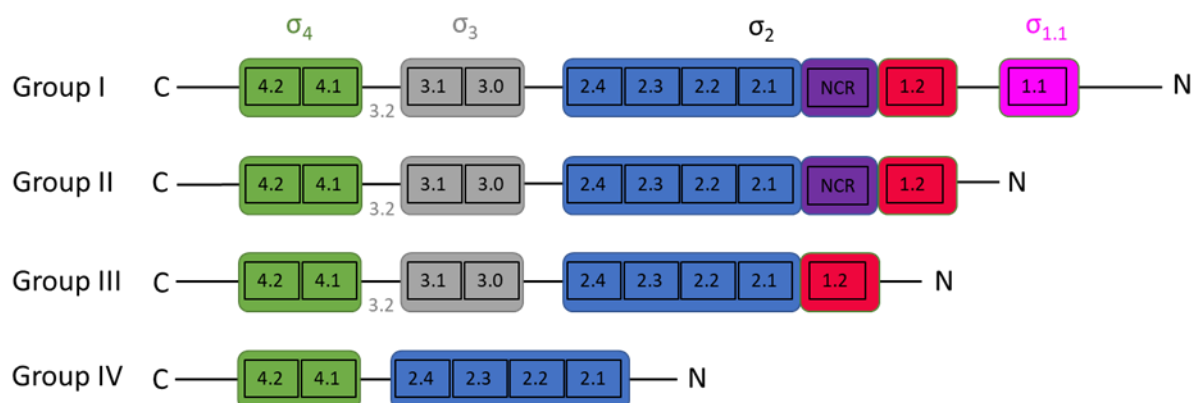


Figure 1.3 The domain organisation for the four sigma factor groups. The four sigma sub-groups of σ^{70} family. The σ group I includes domain $\sigma_{1.1}$ in fuchsia, σ_2 in blue/purple and red, σ_3 in grey and σ_4 in green. The σ group II is structurally the same as group I. The σ group III does not include domain $\sigma_{1.1}$. Group IV does not include domains σ_3 and $\sigma_{1.1}$. (This figure is modified from Paget., 2015).

Group II of the σ^{70} family are structurally similar to group I but lack domain $\sigma_{1.1}$ (Fig. 1.4) and are not required for the transcription of essential growth-related genes (Paget, 2015). This group is known to be capable of recognising the same promoter -10 and -35 elements as the group I sigma factors (Lonetto et al., 1992). The well characterised σ^5 , also known as KatF or

σ^{38} in *E. coli* is classified as a group II sigma factor. During the transition of *E. coli* from exponential to stationary phase, σ^s complexes with RNA polymerase and activates the transcription at the promoters of >500 genes or operons in response to a wide range of stresses, including nutrient limitation, pH alterations, osmotic changes and DNA damage (Lange & Hengge-Aronis, 1994). Almost 140 of the σ^s -dependant genes share the promoter sequence motif 5'-TCTATACTTAA-3' (Weber et al., 2005) and holoenzyme composed of σ^s or σ^{70} can activate the same promoters *in vitro* (Kusano et al., 1996), subtle changes in the extended -10 regions can sway promoter preference to either holoenzyme with σ^s or σ^{70} .

Group III of the σ^{70} family are structurally and functionally diverse from group I and II sigma factors and usually contains sigma domain, σ_2 , σ_3 and σ_4 but lacks domain $\sigma_{1.1}$ (Fig. 1.4). Group III sigma factors are implicated in a general stress response, sporulation, heat shock and flagellar biosynthesis (Paget & Helmann, 2003; Wösten, 1998).

Group IV only consist of the structural sigma domains σ_2 and σ_4 , and lack domains $\sigma_{1.1}$, $\sigma_{1.2}$, σ_3 (Fig. 1.4). This group is referred to as the Extra-Cytoplasmic Factors (ECF) and are involved in sensing and responding to signals generated outside of the cell or in the membrane. The lack of σ_3 makes the ECFs more dependent on the σ_4 to bind to the -35-promoter element (Campagne et al., 2014). In addition to this, the group I σ_2 domain diverges from group IV σ_2 in the $\sigma_{2.3}$ domain which means that the recognition of the -10 element is slightly different in group IV (Campagne et al., 2014). An example of a group IV sigma factor in *E. coli* is σ^E , associated with responding to cell envelope stress. The σ_2 domain of σ^E binds to the non-template ssDNA (single stranded DNA) at the -10 promoter element and has specificity for the motif, 5'-G₋₁₂ TCNNN-7 -3'. Unlike group I sigma factors which require A₋₁₁ and T₋₇ at the -10 promoter element to flip out of the dsDNA, the sigma factor, $\sigma^E_{2.3}$ only requires C₋₁₀ to be flipped out (Campagne et al., 2014).

1.1.5.1 The $\sigma_{1.1}$ region

The N-terminal domain of σ^{70} can be divided into sub-domains 1.1 and 1.2 (Wilson & Dombroski, 1997). The $\sigma_{1.1}$ domain is made up of three-helix bundles, connected to $\sigma_{1.2}$ domain by a flexible linker. The $\sigma_{1.1}$ domain is highly conserved across only group I sigma factors and does not contact DNA but assists in the conversion of RNA polymerase closed complex (RPC)

to RNA polymerase open complex (RPO; Wilson & Dombroski, 1997). The $\sigma_{1.1}$ and a spacer found immediately downstream of domain 1 play an autoinhibitory function, by preventing DNA binding to free σ (Dombroski et al., 1993). In an autoinhibited state, $\sigma_{1.1}$ may crosslink to σ_4 , thus preventing binding of σ_4 to the -35-promoter sequence (Schwartz et al., 2008). The $\sigma_{1.2}$ domain is conserved across primary and alternative sigma factors (Dombroski et al., 1993), and is important for the overall function of σ^{70} .

1.1.5.2 The σ_2 domain

The binding of the -10 promoter sequence to σ_2 , causes a 90° bend in the downstream DNA backbone, which makes contact with the active site cleft of RNA polymerase (Feklistov, 2013). The active site cleft can only accommodate ssDNA and therefore unwinds the dsDNA (Zhang et al., 2012). The σ_2 domain binds to the single stranded -10 consensus (T₋₁₂ A₋₁₁ T₋₁₀ A₋₉ A₋₈ T₋₇) promoter element and stabilises the separated DNA strands (Feklistov & Darst., 2011; Fig. 1.5), this domain is not essential during exponential growth. The $\sigma_{2.3}$ domain recognises the bases of the ssDNA non-template DNA at the -10 element. The highly conserved DNA bases of the non-template ssDNA at A₋₁₁ and T₋₇ flip out of the double helix stack and buries itself in a hydrophobic pocket of σ_2 (Fig. 1.5; Marr & Roberts, 1997). This event causes a process of isomerisation from the RPc to RPO, whereby ~13 bp upstream of the -10 promoter region is melted to expose the template DNA (Gries et al., 2010). The $\sigma_{2.4}$ region recognises dsDNA upstream of the -10 element and contacts the base at position -12, which supports the flipping out of A₋₁₁. In *E. coli*, the W-433 residue which lies within σ_2 is implicated in the stabilisation of the dsDNA and ssDNA transcription bubble junction.

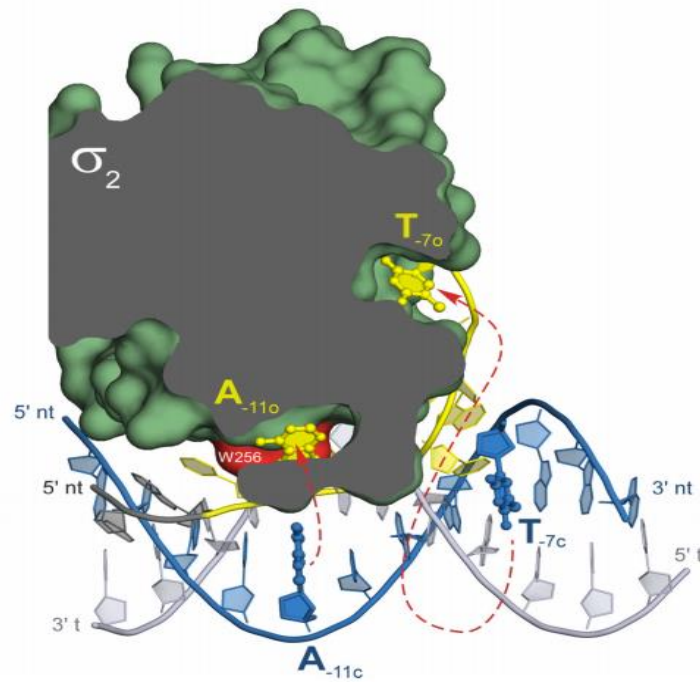


Figure 1.4. Sliced view of the σ_2 domain binding to promoter DNA in *Thermus aquaticus*. This is a magnified view of RNA polymerase holoenzyme in *T. aquaticus*. The σ_2 domain is bound to the -10 promoter double helix in blue which represents RPc, whilst the yellow bases A₋₁₁ and T₋₇ of the non-template strand in yellow represents RPo. This figure was modified with permission from Feklistov and Darst, 2011 (Elsevier).

1.1.5.3 The σ_3 domain

The σ_3 domain is composed of three α helices, sub-domain $\sigma_{3.0}$ is responsible for binding to the major groove of the extended -10 promoter element and the stabilisation of the RPo allowing the DNA to bend towards the σ_3 domain (Zuo & Steitz, 2015). The $\sigma_{3.2}$ loop lies within the active site of RNA polymerase and it has been suggested that the clashing between the growing nascent RNA strand and $\sigma_{3.2}$ loop, causes abortive initial RNA synthesis (Samanta & Martin, 2013). The $\sigma_{3.2}$ loop is not believed to make direct contact with the initiating nucleotides but rather stimulates the binding of the initiating nucleotides, $\sigma_{3.2}$ mutants are still able to stabilise small RNAs but the production of long transcripts prevents promoter escape (Kulbachinskiy & Mustaev, 2006).

1.1.5.4 The σ_4 domain

$\sigma_{4.2}$ region is composed of a helix-turn-helix (HTH) which recognises the -35 promoter element, TTGACA, and inserts into the non-template strand and major groove of DNA, causing 30° DNA bending (Zuo & Steitz, 2015). Promoter escape involves the release of the σ_4 domain from the β flap which destabilises interactions with -35 element (Murakami & Darst, 2003).

1.1.5.5 Structure and function of anti-sigma factors

The activity of some alternative sigma factors is partially controlled by anti- σ factors and adaptor proteins (Treviño-Quintanilla et al., 2013). Anti-sigma factors often consist of a σ -binding domain and sensor domain that responds to signals received from inside or outside of the cell (Paget, 2015). In general, anti- σ factors prevent σ factors from associating with RNAP by inserting between the σ_2 and σ_4 domains and/or wrapping around the compact σ_2 and σ_4 domains. The anti- σ factor is released from the σ factor by regulated proteolysis, partner-switching or conformational change brought about by direct sensing (Paget, 2015). An example of regulated proteolysis is of σ^E /RseA in *E. coli*, which responds to outer membrane dysfunction. An accumulation of unfolded outer membrane proteins in the periplasm causes the C-terminus of RseA to undergo cleavage mediated by a periplasmic protease, DegS (Flynn et al., 2004). This proteolysis activates another protease, RseP which cleaves the cytoplasmic RseA domain and generates a ClpXP recognition tag which is degraded and subsequently releases σ^E (Flynn et al., 2004).

Some anti- σ factors can directly sense signals and respond by undergoing a conformational change to release the σ factor. An example of this is the σ^R /RsrA system in *S. coelicolor*. RsrA lacks membrane spanning or periplasmic domains but is cysteine-rich whilst σ^R is a global regulator of gene expression which provides protective mechanisms to cope with disulphide stress (Paget et al., 2001). Oxidative stress causes disulphide bridges to form in cytoplasmic proteins. σ^R mutants have an increased sensitivity to diamide and the induction of diamide stress at the hyphal structures of *S. coelicolor* leads to activation of σ^R which initiates the transcription of the *trxBA* operon, this produces thioredoxin, thioredoxin reductase and more σ^R (Paget et al., 1998). It is the induction of diamide stress which causes the formation of intramolecular disulphide bonds in RsrA leading to conformational changes and the release of

σ^R , resulting in binding to the *trxB* operon (Paget et al., 2001). This activates the thioredoxin pathway which reduces RsrA back to a state it can hold σ^R inactive (Paget et al., 2001).

An example of partner switching can be found in *B. subtilis* with σ^B , anti- σ factor (RsbW), anti-anti- σ factor (RsbV) and a phosphatase complex (RsbTU or RsbQP; Paget, 2015). RsbTU and RsbQP consist of a phosphatase component (RsbU/ RsbP) and cognate activator (RsbT/ RsbQ) which recognise the environmental stresses. σ^B is held in an inactive state by RsbW during unstressed conditions. During a general stress response, RsbTU or RsbQP dephosphorylates RsbV, RsbV is free to bind to RsbW, subsequently liberating σ^B (Paget, 2015).

1.1.5.6 *Streptomyces coelicolor* sigma factors

Before the *S. coelicolor* genome sequencing, RNA polymerase was found to exist in multiple forms in *S. coelicolor*, holoenzymes composed of σ^{35} , σ^{28} and σ^{49} (Buttner et al., 1988; Westpheling et al., 1985). A probe designed with 10 amino acids from the σ^A , *rpoD* in *Bacillus subtilis* were used to identify four sigma factors, σ^{HrdA} , σ^{HrdB} , σ^{HrdC} and σ^{HrdD} in *S. coelicolor* (Buttner et al., 1990). The *hrdA*, *hrdC* and *hrdD* mutants do not display any phenotype whilst the failed attempts to disrupt *hrdB* indicated its essential function as a primary sigma factor in *S. coelicolor* (Buttner et al., 1990). Later, the genome sequence of *S. coelicolor* revealed 65 σ factors, 45 of which belong to the ECF family (Bentley et al., 2002).

1.1.5.7 *Mycobacterium tuberculosis* sigma factors

The *M. tuberculosis* genome encodes 13 σ factors which all belong to the σ^{70} family (Cole et al., 1998; Manganelli et al., 2004). σ^A , σ^B and σ^F belong to groups I, II and III respectively whilst the remaining 10 σ factors belong to group IV (Cole et al., 1998). The *M. tuberculosis* group I sigma factor, σ^A is constitutively expressed and mRNA expression remains the same despite a multitude of stresses (Manganelli et al., 1999). The expression of *sigA* may increase upon phagocytosis by macrophages in clinical *M. tuberculosis* strains (Rodrigue et al., 2006; Wu et al., 2004). σ^B belongs to the group II sigma factors, and shares homology with conserved regions 2, 3 and 4 of group I (Rodrigue et al., 2006). *sigB* is located 3 kb downstream of *sigA* and is involved in a stress response, $\Delta sigB$ is completely viable in human macrophages

(Doukhan et al., 1995). σ^F belongs to group III and is induced in response to many stresses including, antibiotics, cold shock, oxidative stress, nutrient limitation and during stationary phase.

1.1.6 Transcription initiation

The dissociable sigma subunit binds to RNA polymerase and forms a holoenzyme which dictates specific promoter recognition. The $\sigma_{2.4}$ region binds to the dsDNA at position -12 of the DNA fork junction. The $\sigma_{4.2}$ region recognises position -35 of the non-template DNA, causing the DNA to bend 30° (Zuo & Steitz, 2015; Fig. 1.6A). The promoter DNA lies across one face of the holoenzyme and outside of the RNAP active-site channel (Murakami & Darst, 2003). The $\sigma_{1.1}$ region may play an important role in the RPo complex formation by stabilising promoter-complexes (Miropolskaya et al., 2012). The positioning of the $\sigma_{3.2}$ region in the RNA exit channel suggests that it may be involved in promoter clearance by RNAP, mutations in $\sigma_{3.2}$ are unable to perform abortive transcription or promoter escape (Pupov et al., 2014). During the intermediary step, the melted promoter DNA is temporarily stabilised whilst also introducing flexibility into the DNA (Murakami & Darst, 2003), ready for entry into the active site channel (Fig. 1.6B). The RNAP-promoter open complex, RPo, occurs when upstream DNA bends around RNAP at positions -35 (36°) and -25 (8°). At position -16, the DNA makes another sharp turn at 37° towards RNAP (the DNA melting extends past the transcriptional start site) (Murakami et al., 2002). The non-template ssDNA at the -10 element interacts with the exposed aromatic residues of $\sigma_{2.3}$ domain. Positions -2 to +4 are held in a groove between $\beta 1$ and $\beta 2$ and the ssDNA is directed into the positively charged active site of RNAP by bending 90°, where it pairs up with the initiating nucleotides (Murakami & Darst, 2003). The RNA synthesis result is an RNA molecule, 12 nucleotides long which fills the RNA-DNA hybrid and the upstream RNA exit channel, this displaces the $\sigma_{3.2}$ loop and ends abortive initiation (Fig. 1.6D). The displacement of $\sigma_{3.2}$ causes the destabilisation of interactions between the σ_2 and σ_4 at positions -10 and -35, respectively (Fig. 1.6E). RNAP can then escape transcription initiation into elongation (Fig. 1.6F; Murakami & Darst, 2003).

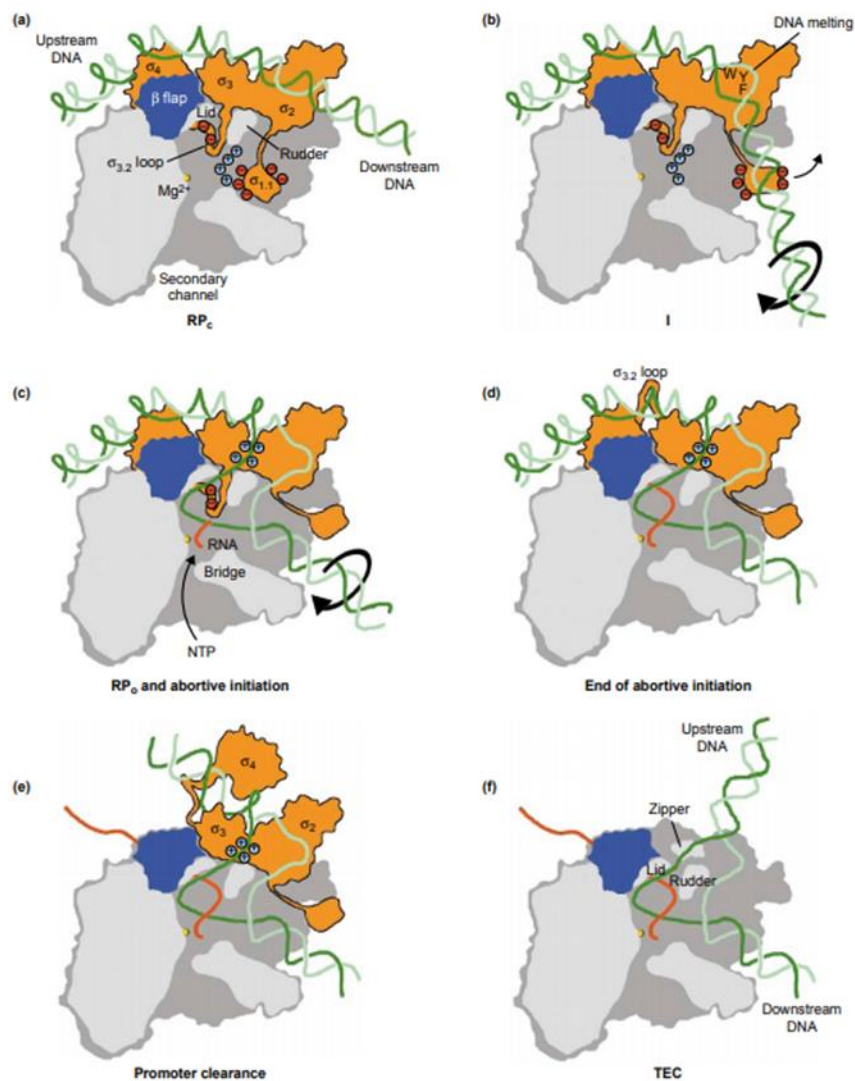


Figure 1.5 The steps of transcription initiation. (A) Recognition of promoter -35 and -10 DNA elements and formation of the initial RP_c. (B), (C) Intermediate and start of RP_o and abortive initiation. (D) End of abortive initiation. (E) The σ dissociates and RNAP undergoes promoter clearance. (F) Transcription elongation complex. Figure received permission from Murakami and Darst, 2003 (Elsevier). σ (orange), non-template DNA (light green), template DNA (dark green), RNAP (subunits in grey and β flap in blue and Mg²⁺ in the active site is yellow) together form the transcription initiation complex.

Transcription initiation is usually the target for many transcription factors that act as activators or repressors of transcription. Some examples of these include: CRP and bacteriophage λ

repressors (Lloyd et al., 2001). The cyclic AMP receptor protein, CRP binds to the C-terminal domain of the RNAP α subunit and modulates the expression of >100s of genes in response to changes in cAMP concentrations (Savery et al., 1998). The CRP binding sites are usually upstream of the -35 promoter element or overlapping the -35 sequence (Savery et al., 1998). Some activators interact with the RNAP σ subunit, others alter the promoter conformation (Lloyd et al., 2001). Repressors tend to bind to promoter DNA or close to the transcriptional start site to prevent RNAP binding. The lac operon includes interactions between RNAP, repressor and operator and is good example of repressor activity. In the absence of the inducer, the lac repressor binds to the operator and prevents transcription of the lac operon (Becker et al., 2013; Schmitz & Galas, 1979)

1.1.6.1 Transcription elongation

The elongation complex is formed directly after RNAP promoter escape, usually with a significant conformational change as a result of the loss of σ (Vassilyev et al., 2007). The largest conformational change occurs at the β and β' RNAP domains (crab-claw-like structure) which bind to σ in the RNAP holoenzyme (Vassilyev et al., 2007). The downstream DNA enters RNAP and approximately 9 nt of RNA at the growing end are annealed to the template DNA strand, forming a RNA/DNA hybrid (Roberts et al., 2008).

RNAP can move backwards along the DNA template, also referred to as 'backtracking', by reversing the steps of the nascent RNA assembly back towards the RNAP secondary channel (Borukhov et al., 2005). Backtracking is a response to errors that occur during transcription such as mis-incorporation or mispairing of the RNA/DNA hybrid (Roberts et al., 2008). The Gre proteins, GreA and GreB in *E. coli* stimulate hydrolysis of the backtracked elongation complex, by cleaving the extruding portion of 3' RNA so that transcription elongation can resume (Borukhov et al., 2005). GreA prevents transcription arrest whilst GreB can reactivate pre-arrested transcription complexes (Borukhov et al., 2005).

The transcription repair coupling factor, Mfd reactivates or recycles stalled or arrested RNAPs, through reverse backtracking, allowing RNAP to re-engage with the 3' RNA. A proposed model suggests, Mfd binds to the stalled RNAP and causes disassociation of RNAP and RNA from the template DNA. Mfd remains slightly attached to RNAP, and complexes with UvrA-UvrB, two

damage recognition proteins, which result in the complete release of Mfd/RNAP and delivery of UvrA-UvrB to the damage site (Selby, 2017). UvrA-UvrB excise the phosphodiester backbone of damaged DNA, five bases at 3' end and eight bases 5' end of the damaged site (Selby, 2017). Mfd increases DNA repair upon UV damage and *mfd* knockouts increase the frequency of mutations (Selby, 2017).

1.1.6.2 Transcription termination

The final stage of transcription is termination, RNAP dissociates from the DNA template and as a result the nascent RNA molecule is released. There are two types of termination in prokaryotes: intrinsic termination (Rho-independent) and Rho-dependent termination.

Transcription factors can bind to elongation complexes and affect/modulate termination in a cell (Ray-Soni et al., 2016). For example, NusA and NusG binds to elongating RNAP and nascent RNA and influences transcriptional pausing for transcription termination and anti-termination. NusA competes with the RNA binding factor, Rho which terminates transcription of untranslated operons, for termination binding sites (Yakhnin et al., 2016; Qayyum et al., 2016). Intrinsic termination occurs at inverted repeats directly followed by multiple uridine residues, with the formation of a stem-loop structure in the nascent RNA (Washburn & Gottesman, 2015). Intrinsic termination is governed by specific timing of the stem loop formation, size, length and sequence before RNAP moves towards the hairpin structure. The transcriptional elongation complex pauses upon encountering the hairpin structure and dissociates from the weak RNA(U) (U-tract)/DNA(A) hybrid at the seventh or eighth uridine, releasing the DNA template (Czyz et al., 2014; Washburn & Gottesman, 2015). There are cases of intrinsic termination occurring without a U-tract in the RNA molecule, however this is not fully understood (Czyz et al., 2014). Unlike intrinsic termination, Rho-dependent termination occurs in the absence of RNA secondary structures. The homohexameric RecA-family helicase, Rho binds to the RNA as it exists through the RNAP exit channel. Rho favours binding to *rut* sites which are C-rich and G-poor, this also reduces the formation of secondary structures at the Rho binding sites. These *rut* sites are found upstream of the termination site, usually 80-90 nt long (Bear et al., 1988), which is sufficient for the binding of the six monomeric Rho proteins, each monomer binds to 13 nt on the RNA molecule (Bear et al., 1988; Ray-Soni et al., 2016).

Upon Rho binding to *rut* RNA, ATP induces ring closure of Rho and activation of catalysis. Rho translocates the nascent RNA via ATP hydrolysis (Ray-Soni et al., 2016).

1.1.7 RNAP binding regulatory factors

Transcriptional activators of the RNAP- σ^{70} holoenzyme increase transcription from promoters by three different mechanisms: class I and class II activate and/or cause a conformational change of promoter DNA (Wu et al., 2018). Class I activation involves an activator binding upstream of the -35 DNA element and recruiting RNAP. Class II activators bind to the binding site overlapping the promoter -35 and recruits RNAP through interactions with the α subunit and σ_4 . Lastly, some proteins can bind to -35 and -10 promoter DNA and alter the DNA structure which promotes RNAP binding (Wu et al., 2018). Aside from sigma factors interacting with RNAP to regulate transcription, there are other proteins that also regulate the transcription cycle. For example, the Gre factors (1.2.1.11), GreA and GreB resolve backtracked complexes by interacting with the secondary channel of RNAP and repositions the 3' nascent RNA into the active cleft of RNAP and encourages promoter escape (Washburn & Gottesman, 2015).

1.1.7.1 ppGpp

The response to starvation is referred to as the stringent response and causes extensive transcriptional reprogramming (Ross et al., 2013). The stringent factor, ppGpp is responsible for the detection and response to nutrient starvation (Hesketh et al., 2007). ppGpp regulates transcription in *E. coli* by binding to two allosteric sites of RNAP at ppGpp-specific promoters (Ross et al., 2013). Additionally, ppGpp inhibits DNA replication by binding to the active site of RNAP primase DnaG (Maciag et al., 2010). The level of (p)ppGpp is balanced by pyrophosphokinase and pyrophosphohydrolase activities of, two homologous enzymes in *E. coli*, SpoT and RelA. Under amino acid limitation, the *E. coli*, RelA is associated with ribosomes and is responsible for ribosome-dependent synthesis of ppGpp as it binds to the uncharged tRNA (Potrykus et al., 2011). The stringent response also occurs in *Streptomyces* spp and is thought to be important in the triggering of antibiotic production in certain strains. The *S.*

coelicolor homologue $\Delta relA$ strain demonstrates a deficiency in ACT and RED (Chakraborty & Bibb, 1997; Sun et al., 2001). SpoT-dependent ppGpp synthesis is activated by carbon and phosphate limitation in a ribosome-independent manner (Sun et al., 2001). In *S. coelicolor*, transcription of *rshA* is activated upon ppGpp synthesis and over-expression of *rshA* can restore antibiotic production and sporulation (Sun et al., 2001). Both *relA* and *rshA* are similar to *E. coli* SpoT as they display hydrolase and synthetase activity (Sun et al., 2001). Yet the mechanism of ppGpp action remains unclear in *Streptomyces* but is unlikely similar to the ppGpp function in *E. coli* as the binding sites for RNAP are absent.

1.1.7.2 Crl

Crl was originally found in the cytoplasm of *E. coli* and was thought to activate the cryptic curlin subunit gene which results in curli formation and fibronectin binding (Arnqvist et al., 1992). However, more recent work shows Crl binds to the master regulator of general stress, σ^S and positively regulates σ^S -dependant genes (Monteil et al., 2010). Crl can bind to free σ^S and the RNAP/ σ^S complex (Gaal et al., 2006). Crl has the ability to also bind to σ^{70} at lower binding affinities, but more likely increases holoenzyme formation with σ^S (Gaal et al., 2006). Most importantly, unlike many other transcriptional regulator, Crl does not bind to DNA (Gaal et al., 2006).

1.1.7.3 DksA

DksA (DnaK suppressor A) is a 17.5 kDa protein conserved in proteobacteria and is composed of five α helices organised into three domains (Perederina et al., 2004). DksA modulates RNAP activity in a similar way to GreA through insertion of a coiled coil into the secondary channel of RNAP (Wilma Ross et al., 2016). Both DksA and ppGpp act as important regulators in the stringent response and together can negatively and positively regulate the expression of many genes (Perederina et al., 2015). DksA and ppGpp together activate amino acid biosynthetic promoters and stable RNA promoters (Paul et al., 2005) and reduce the life time of RNAP/promoter open complexes, which can either be advantageous or disadvantageous

depending on the promoter (Molodtsov et al., 2018). Aside from transcription initiation, DksA has been identified as an elongation factor (Zhang et al., 2014). ChIP experiments reveal that DksA co-localises at both promoters for transcription initiation (with σ^{70} and RNAP) and downstream regions which eluded to a possible function during elongation (Zhang et al., 2014). Amino acid starvation causes stalling of RNAP at mRNA by uncoupling of transcription and translation, promoting backtracking of RNA and DNA through RNAP (Zhang et al., 2014). DksA prevents backtracking by arresting elongation complexes *in vivo* (Zhang et al., 2014).

1.1.7.4 RapA (HepA)

RapA is 110 kDa protein which consistently co-eluted with RNAP preparations in *E. coli* and is as abundant as σ^{70} . This contaminating protein is the product of the *rapA* gene and was classified as a member of the SNF2 family of putative helicases (Muzzin et al., 1998). The early *in vitro* experiments showed RapA binds to core RNAP but not holoenzyme, this was later followed up and showed RapA and σ^{70} compete for RNAP. Members of the SNF2 family are associated with DNA repair processes, therefore upon UV exposure, $\Delta rapA$ demonstrated reduced survival (Muzzin et al., 1998) and inhibits growth at high salt concentrations (Sukhodolets et al., 2001). RapA shares high sequence similarity to ATPases and the solved crystal structure of RapA confirmed an ATP binding domain and dsDNA-binding site (Shaw et al., 2008). RapA is believed to play a role in back translocation of transcription, backtranslocation can have a negative impact on genome stability, however RapA is believed to have low affinity for DNA and can easily be dissociated from DNA by RNAP during transcription in the forward direction (Liu et al., 2015).

1.1.7.5 RNA polymerase binding protein, RbpA

RbpA is a 14 kDa RNA polymerase binding dimeric protein that is conserved across the Actinobacteria phylum (Fig. 1.7) and the was first detected through consistent co-elution with *S. coelicolor* RNAP during affinity chromatography and gel filtration (Newell et al., 2006; Paget et al., 2001). Additionally, ChIP-qPCR has shown the presence of RbpA in a RNAP- σ^{HrdB} /promoter- complex (Tabib-Salazar et al., 2013) and is essential for the growth of *S.*

coelicolor and *M. tuberculosis* (Hu et al., 2014). RbpA (RbpA^{Sc} and RbpA^{Mtb}) was found to activate transcription from promoters that are recognised by the principal sigma factors, σ^{HrdB} or the orthologue σ^A in *M. tuberculosis* (Newell et al., 2006; Tabib-Salazar et al., 2013). Structural studies have shown RbpA binds specifically to the σ_2 domain and the β subunit of RNAP to initiate transcription (Tabib-Salazar et al., 2013).

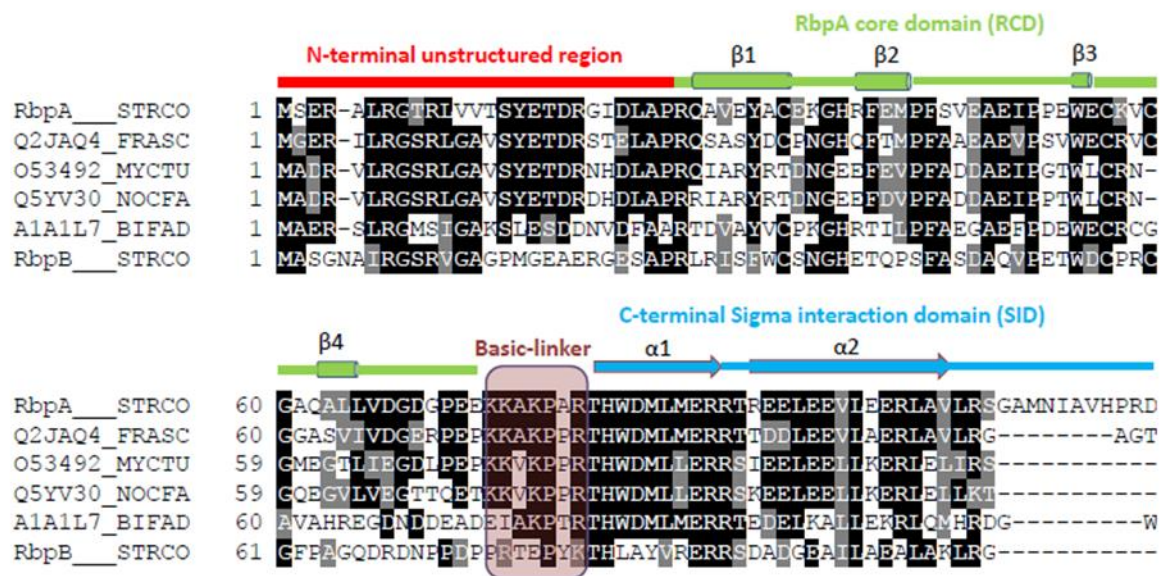


Figure 1.6. Multiple sequence alignment of selected RbpA homologues. ClustalW analysis of the amino acid sequences from selected RbpA homologues. The RbpA core domain (RCD) is highlighted green and shows all four of the β strands, whilst the C-terminal Sigma interaction domain (SID) is highlighted blue with both α -helices annotated. Residues with a black box represent identical whilst the grey box indicates similar amino acids amongst the organisms. STRCO-*Streptomyces coelicolor*, FRASC-*Frankia spp*, MYCTU-*Mycobacterium tuberculosis*, NOCFA-*Nocardia farcinica*, BIFAD-*Bifidobacterium adolescentis*.

1.1.7.5.1 Structure of RbpA

The RbpA N-terminus is an unstructured 26 amino acid tail that is highly conserved within the Actinobacteria, and the function of this region remains largely unknown. A cryo-EM structure of the RbpA-NTT (N-terminal tail) has recently been solved, positioning RbpA-NTT into the RNAP active site cleft between β zinc-binding domain (ZBD) and σ^A_4 domains and snakes towards the RNAP active cleft Mg^{2+} (Boyaci et al., 2018; Fig. 1.8). The RbpA-NTT interacts with

the $\sigma_{3.2}$ and the RNAP ZBD and β' lid. The interactions between RbpA-NTT and the σ -finger suggests that it plays a role in RNAP initiation, although not in RPo formation (Boyaci et al., 2018). Deletion of the RbpA-NTT does not appear to have deleterious effects on the phenotype (Prusa et al., 2018), however RNA-sequencing has shown changes in the transcriptome, strongly suggestive that the NTT may be important for regulation of specific promoters (Hubin et al., 2017a). The designated RbpA core domain (RCD) forms a β barrel, which is composed of four β strands (Tabib-Salazar et al., 2013). The C-terminus comprises of two α -helices considered to be involved in the binding of σ^{HrdB} or σ^{HrdA} in *S. coelicolor*, and σ^{A} or σ^{B} in *M. tuberculosis*; it is therefore designated the sigma interaction domain (SID). Complete removal of the C-terminal domain renders RbpA incapable of binding the σ factor (Tabib-Salazar et al., 2013). In the absence of core RNAP, RbpA is still capable of binding to the σ_2 domain of the principal σ factor, HrdB (Bortoluzzi et al., 2013; Hu et al., 2014). Two critical arginine residues, R89 and R90 are essential for σ binding and are well conserved across the Actinobacteria (Prusa et al., 2018; Tabib-Salazar et al., 2013). Between the RCD and SID exists a 15-amino acid lysine-rich basic linker (BL) which consists of four positively charged amino acids K73, K74, K76 and R79 in *M. tuberculosis* (K74, K75, K77 and R80 in *S. coelicolor*). RbpA^{R79} in *M. tuberculosis* is positioned to interact with -13 and -14 negatively charged non-template DNA whilst RbpA^{K73}, RbpA^{K74} and RbpA^{K76} are positioned to make long-range electrostatic interactions with -18/-19 and -17/-18 of the template DNA (Hubin et al., 2015; Hubin, et al., 2017a). The RbpA^{R79} mutation proved to be important for the activation of transcription *in vitro*.

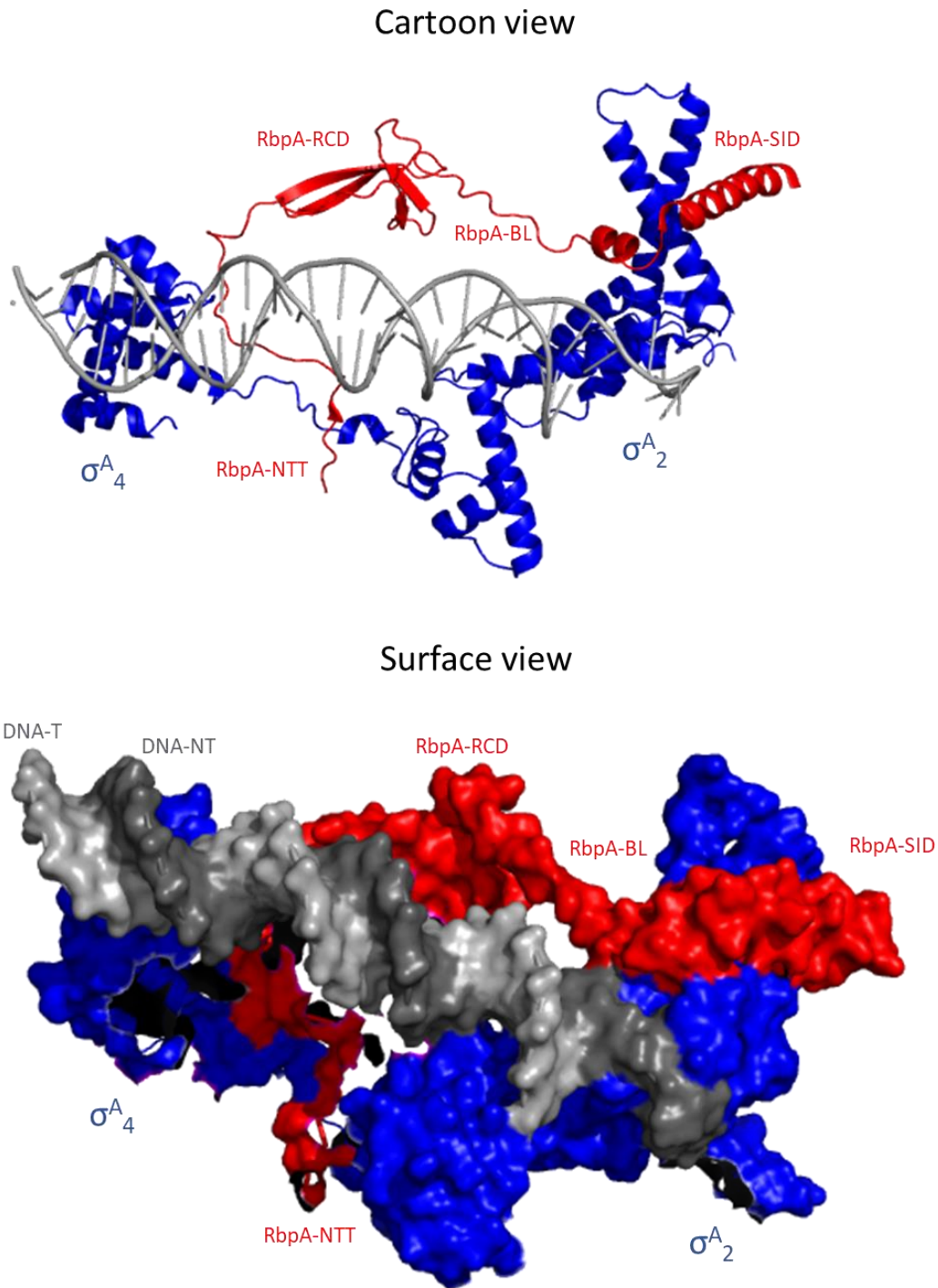


Figure 1.7. The cryo-EM structure of *M. tuberculosis* σ^A /RbpA/up-stream fork DNA. The non-template DNA strand is shown in dark grey, template DNA in light grey, RbpA in red with the N-terminal tail (NTT), RbpA core domain (RCD), basic linker (BL) and sigma-interaction domain (SID) annotated. Structure downloaded from RCSB PDB, accession number 6C04 and visualised with PyMOL (version 2.0.7; Boyaci et al., 2018).

1.1.7.5.2 Functions associated with RbpA

Many of the initial studies on RbpA eluded to a stabilisation role of the transcription initiation bubble and holoenzyme complex or activation of transcription (Tabib-Salazar et al., 2013). This has been assessed *in vitro* by testing the stability of RNAP- σ^A holoenzyme and RbpA increases the retention of σ^A and σ^B at the holoenzyme (Hu et al., 2014). The holoenzyme complex without RbpA was unable to form stable promoter complexes and thus unable to initiate transcription (Hu et al., 2014). RbpA is also important for the stabilisation of the RPo complex and mutations that affect DNA and σ binding do not affect overall promoter opening driven by RbpA (Prusa et al., 2018). RbpA has also been shown to rescue transcription activation from promoters with a poorly conserved -35 promoter DNA (Hubin et al., 2017). RbpA is induced by disulphide stress and rifampicin treatment whilst the $\Delta rbpA$ exhibits 15-fold increased sensitivity to rifampicin (Newell et al., 2006). Many theories were suggested for the increased sensitivity to rifampicin, firstly RbpA may directly compete with rifampicin for the RNAP binding site, secondly RbpA may change the binding site of rifampicin-RNAP and thirdly, RbpA may affect downstream processes by changing expression of cell proliferation genes (Hu et al., 2014). However, another study shows that RbpA can rescue the *in vitro* activity of RNAP even in the presence of 100 μ M rifampicin (Dey et al., 2010).

Interestingly, a study has shown that *M. tuberculosis* can produce a small percentage of INH-tolerant cells in response to INH treatment, and σ^B is important for tolerance to several drugs (Wang et al., 2018). RbpA can bind to group II sigma factors, such as σ^B , deletion of σ^B does not change the MIC but does decrease the fraction of cells tolerant to INH. The σ^B -dependant gene, *ppk1*, responsible for polyP biosynthesis is proportionate to σ^B levels. The RbpA- σ^B -*ppk1* complex is believed to be important for INH tolerance in *M. tuberculosis*, any deletions of the key players of this complex reduce the number of cells tolerant to INH (Wang et al., 2018). Additionally, an antimicrobial RNAP inhibitor, Fidaxomicin (Fdx) used for the treatment of *Clostridium difficile* infection has shown activity against multi-drug resistant *M. tuberculosis* RNAP *in vitro* (Boyaci et al., 2018). Cryo-EM structures have revealed that the binding site of Fdx is independent to the rifampicin binding site on RNAP and involves interactions with RNAP β' - σ^A -RbpA-NTT (Boyaci et al., 2018). The RbpA-NTT has been shown to play a role in high sensitivity of *M. tuberculosis* cells to Fdx.

1.1.7.6 RNAP binding protein, CarD

CarD is a well-conserved RNAP-binding protein originally described in *Mycobacterium* in 2009 and is present in a wide range of Gram-positive and Gram-negative bacteria (Fig. 1.9). CarD is not found in *E. coli* and it has not been investigated in any other Actinobacteria (Rammohan et al., 2015). Attempts to delete *carD* in *M. tuberculosis* and *M. smegmatis* were unsuccessful suggesting that it is essential for viability (Stallings et al., 2009).

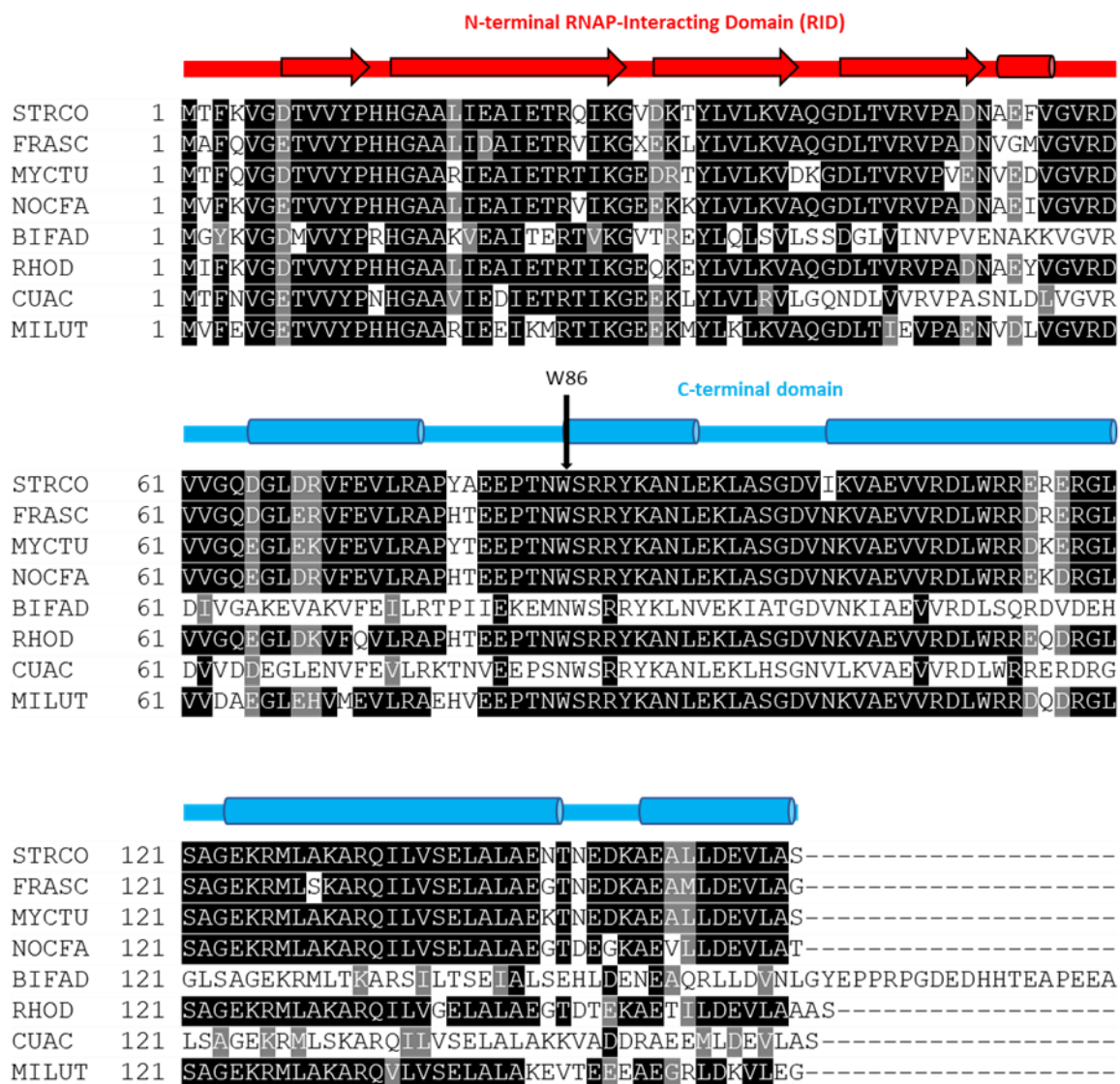


Figure 1.8 Multiple sequence alignment of CarD homologues. A multiple sequence alignment of the CarD homologues from the actinobacteria phylum, performed using BLASTP to look at similar sequences and clustalW, a multiple sequence alignment tool. BOX shading analysis was performed using BOXSHADE (version 3.21). Identical amino acid residues are shaded in black

and similar amino acids are shaded grey. The RNA polymerase interacting domain (Red) and the C-terminal domain (blue). β strands as arrows and α helices as cylinders. The amino acids highlighted black are identical in each homologue, and grey represents similar amino acids. STRCO= *Streptomyces coelicolor*, FRASC= *Frankia* spp, MYCTU= *Mycobacterium tuberculosis*, NOCFA= *Nocardia farcinica*, BIFAD= *Bifidobacterium adolescentis*, RHOD= *Rhodococcus* spp, CUAC= *Cutibacterium acnes*, MILUT= *Micrococcus luteus*

1.1.7.6.1 Structure of CarD

CarD is composed of two domains, the Tudor-like fold of the N-terminal domain known as the RNAP interaction domain (RID) and the C-terminal domain (CTD) (Srivastava et al., 2013). More recently synchrotron-based small angle X-ray scattering experiments suggest that CarD exists in two states, as a oligomer in solution and as a homodimer in solution at higher μ M concentrations (Kaur et al., 2018). Bacterial two-hybrid assays confirmed that the N-terminus of CarD in *T. thermophilus* interacts with the β 1 lobe of RNAP (Garner et al., 2017; Stallings et al., 2009; Fig. 1.10). The N-terminus of CarD shares sequence and RNAP binding similarity to the RID of the transcription-repair coupling factor (TRCF) which is encoded by *mfd* (Srivastava et al., 2013; Stallings et al., 2009). The CarD-CTD has no sequence or structural similarity to any other previously characterised proteins. The CTD is composed of five alpha helices which appear to make direct contact with position -10 of promoter DNA on the opposite face of RNAP as σ (Srivastava et al., 2013).

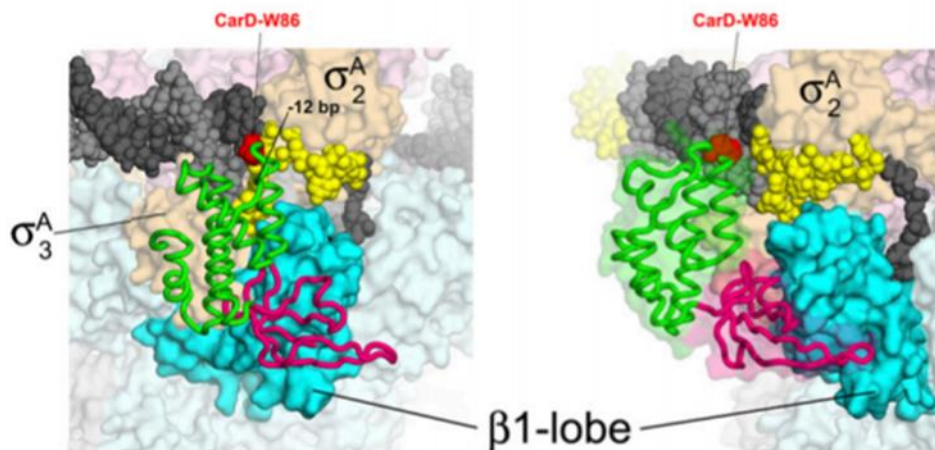


Figure 1.9. Structural model of the RPo/CarD complex in *T. thermophilus*. Molecular surfaces shown for RNAP holoenzyme. CarD-RID (pink), CarD-CTD (green) and W86 (Red). DNA non-template strand (dark grey), template strand (light grey) and promoter positions -35 and -10 (yellow). RNAP β 1 lobe (cyan), β (light blue), σ^A (gold). Left and right images are two different views of the complex. Permission: creative commons licence (Srivastava et al., 2013).

1.1.7.6.2 Functions associated with CarD

CarD is up-regulated (~ 3 to 20-fold) in response to double strand breaks induced by treatment with Ciprofloxacin, Bleomycin and alkylating agent methyl methanesulphonate and nutrient stress (Stallings et al., 2009). This suggested that CarD is necessary for survival during oxidative, DNA damage and nutrient limitation. Whole genome transcriptional profiling indicated that in the absence of CarD, 193 genes were upregulated >2 fold whilst 176 genes were downregulated >2 fold (Stallings et al., 2009). The most significant up-regulation in the absence CarD was the up-regulation of rRNA levels, therefore CarD may play a role in the transcription of rRNA operons during stress conditions. CarD is proposed to play a role in *M. tuberculosis* persistence and bacterial replication in vivo. Mice infected with a wild-type and conditional doxycycline-dependant CarD knock-down strain received doxycycline at day 1 and 56. The *carD*-depletion infections were unable to replicate even after day 1 compared to the wild-type strain infection (Stallings et al., 2009). Most importantly, CarD modulates transcription through direct contact with DNA. ChIP experiments reveal the failure of CarD to bind across the genome in the absence of RNAP and co-localised at most promoters with σ^A (Srivastava et al., 2013). The CTD includes a well conserved tryptophan, W86 in *M. tuberculosis* (W85 in *S. coelicolor*)

surrounded by a basic patch which is reportedly important for binding to DNA. CarD is predicted to interact around the dsDNA (-12) -ssDNA (-11) interface of a transcription bubble, at positions -11 to -15. As a result, W86 interacts with the minor groove of the dsDNA and the attraction is further supported by the basic patch (Srivastava et al., 2013). CarD can bind to RNAP *in vivo* and activate transcription *in vitro*, despite the alanine substitution of W85. Many kinetic studies have also shown CarD has been elucidated in stabilisation of open complexes like RbpA (Davis et al., 2015; Rammohan et al., 2015). Additionally, CarD has been the first cytosolic transcription factor in *M. tuberculosis* to produce amyloid-like fibrils *in vivo* and *in vitro*, whilst the function of this still remains unknown (Kaur et al., 2018).

1.1.7.7 Relationship between RbpA and CarD

We now understand that RbpA (Dey et al., 2010) and CarD (Stallings et al., 2009) are both essential in *S. coelicolor* and *M. tuberculosis*. RbpA and CarD bind to distinct regions of RNA polymerase however both form RNAP-holoenzyme complexes and are both known to stabilise RP₀ individually (Rammohan et al., 2016). Structural modelling suggested that CarD contacts DNA from the opposite side of the DNA double helix to RbpA and there are no steric clashes (Hubin et al., 2015). Recently, an X-ray crystal structure of RbpA-TIC complex has superimposed CarD onto the structure to show the proximity of both transcription factors (Fig. 1.11). Since CarD and RbpA are found at all RNAP- σ^A transcription initiation complexes, it could suggest an overlapping function between both. Disrupting the RbpA/CarD/RNAP complex could be an important source of drug discovery. The *M. tuberculosis* RNAP open complex is less stable compared to a fast-growing organism such as *E. coli* and therefore the combination of RbpA and CarD provides better stability of transcription initiation bubbles and enables transcription activation (Garner et al., 2017; Hu et al., 2012; Rammohan et al., 2016).

Fluorescence assays to investigate the transcription open complexes using a Cy3-labeled promoter showed that the fluorescence fold change of RbpA and CarD is lower individually compared to the presence of both transcription factors in the holoenzyme (Rammohan et al., 2016). Additionally, a study investigated abortive transcription using two well characterised *M. tuberculosis* promoters, *vapB10p* and *rrnAP3*. Full length RbpA and truncations excluding the N-terminal tail were capable of activating transcription of the *vapB10p in vitro* by 2-fold, CarD

alone also increased transcription by 3-fold. The addition of RbpA to *M. bovis* RNAP holoenzyme complexed with CarD activated transcription by 6-fold (Hubin et al., 2017).

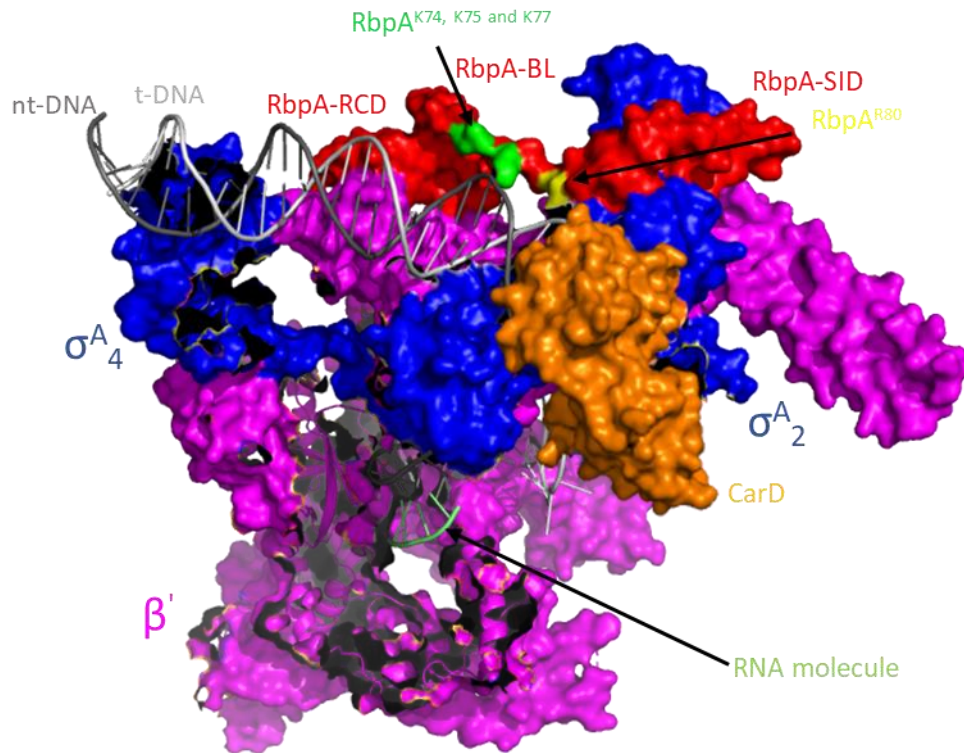


Figure 1.10. X-ray crystal structure of RbpA-transcription initiation complex with superimposed CarD. RbpA-TIC complex in *M. smegmatis*, PDB accession ID: 5VI5 (Hubin et al., 2017b) and superimposed CarD in *T. aquaticus* (Bae et al., 2015), PDB accession ID: 4XLR. RbpA shown in red (basic linker amino acids highlighted in green and yellow), RNAP β' subunit is in magenta, σ^A in blue, CarD in orange, non-template DNA in dark grey, template DNA in light grey and the 4 mer RNA molecule in lime green.

1.1.8 Project aims

The focus of this project was to gain a better understanding of the mechanism of action and biological importance of RbpA and CarD in Actinobacteria, *S. coelicolor* and *M. tuberculosis*. The aims of this thesis are outlined below:

- 1) Understand the biological importance of the RbpA basic linker in *S. coelicolor*
- 2) Understand the biological importance of the RbpA basic linker in *M. tuberculosis*
- 3) Investigate the mechanism and biological importance of *rbpA* in *S. coelicolor*
- 4) To develop a co-depletion strain of both *rbpA* and *carD* to investigate a possible relationship between the two *in vivo*

Chapter 2: Materials and Methods

2.1 Chemicals and reagents

- Agar (Sigma-Aldrich)
- Agarose (Melford)
- Ammonium Persulfate (APS; Sigma-Aldrich)
- Ampicillin (Melford Laboratories)
- Antifoam (Sigma-Aldrich)
- Apramycin (Duchefa Biochemie)
- Bromophenol Blue (Amersham Biosciences)
- Casamino Acids (Difco)
- Chloramphenicol (Melford Laboratories)
- Chloroform (Sigma-Aldrich)
- D (+) maltose monohydrate (Fisher Scientific)
- Deoxyribonucleotide phosphates (dNTPs; New England Biolabs)
- Diamide (Sigma-Aldrich)
- Dimethylsulphoxide (DMSO; Sigma-Aldrich)
- Dithiothreitol (DTT; Melford Laboratories)
- GelRed (Biotin)
- Glycerol (Fisher Scientific)
- HEPES (Sigma-Aldrich)
- Iodoacetamide (Sigma-Aldrich)
- Isopropyl- β -D-thiogalactopyranoside (IPTG; Melford Laboratories)
- Kanamycin (Melford Laboratories)
- Malt extract (Oxoid)
- Mannitol (Sigma-Aldrich)
- MOPS (Fisher Scientific)
- Nalidixic acid (Duchefa Biochemie)
- N, N-dimethyl-formamide (Fisher Scientific)
- Nutrient agar (Difco)
- Peptone (Bacto)
- Phenol (Fisher Scientific)
- Phenylmethylsulfonyl Fluoride (PMSF; Sigma-Aldrich)

- Polyethylene glycol 6000 (Sigma-Aldrich)
- Polyethylene glycol 8000 (Sigma-Aldrich)
- Ponceau C solution (Sigma-Aldrich)
- Protease Inhibitor Cocktail tablets (Sigma-Aldrich)
- Protein A magnetic beads (New England Biolabs)
- Protein G magnetic beads (New England Biolabs)
- P-40 substitute (Calbiochem)
- Rifampicin (Sigma-Aldrich)
- Sigmacote® (Sigma-Aldrich)
- Sodium chloride (Fisher Scientific)
- Sodium hypochlorite (Sigma-Aldrich)
- Sodium deoxycholate (Sigma-Aldrich)
- Sodium dodecyl sulphate (SDS; Fisher Scientific)
- Soya flour (Infinity Foods)
- Sucrose (Sigma-Aldrich)
- TES (N-Tris(hydroxymethyl)methyl-2-aminoethane sulfonic acid; Fisher Scientific)
- Tetramethyl-ethylenediamine (TEMED; Fisher Scientific)
- Thiostrepton (Sigma)
- Trichloroacetic acid (TCA) (Sigma)
- Tris (2-Amino-2-hydroxymethyl-propane-1,3-diol; Fisher Scientific)
- Triton X-100 (Sigma-Aldrich)
- Tryptone (Difco)
- Tween 20 (Sigma-Aldrich)
- Ultra Pure Sequagel® Urea Gel™ Concentrator (national diagnostics)
- Ultra Pure Sequagel® Urea Gel™ gel diluent (national diagnostics)
- Ultra Pure Sequagel® Urea Gel™ buffer (national diagnostics)
- X-gal (Melford Laboratories)
- Yeast extract (Difco/Oxoid)
- 7H9
- 7H10

DNA polymerases

- Phusion High-Fidelity DNA Polymerase (New England Biolabs)
- Taq polymerase (New England Biolabs)
- GoTaq DNA Polymerase (qPCR Master Mix; Promega)

DNA modifying enzymes

- Antarctic Phosphatase (New England Biolabs)
- T4 DNA Kinase (New England Biolabs)
- T4 DNA Ligase (New England Biolabs)
- DNase/RNase free herring sperm DNA (Promega)

DNA/RNA restriction enzymes

- DNA endonucleases (New England Biolabs)
- S1 nuclease (Life Technologies)
- RNase A from bovine pancreas (Sigma-Aldrich)
- RiboShredder™ RNase Blend (Cambio)

2.1.1 Antibodies

- anti-FLAG M2 monoclonal antibody, from mouse, Sigma F18041MG.
- anti- σ^{HrdB} polyclonal antibody, from rabbit, a gift from P. Doughty.
- anti-RNAP β subunit monoclonal antibody, from mouse, Abcam ab12087.
- Anti-SigE polyclonal antibody, from rabbit, a gift from M. Hutchings.

2.1.2 Buffers and solutions

DNA Manipulation

TE buffer (pH 8.0): 10 mM Tris-HCl (pH 8.0), 1 mM EDTA

10X TAE electrophoresis buffer (1 L): 48.4 g Tris base, 11.4 mL glacial acetic acid, 3.7 g EDTA, dissolve in 1 L dH₂O

10X TBE electrophoresis buffer (1 L): 108 g Tris base, 55 g boric acid, 9.3 g Na₂EDTA, dissolved in 1 L dH₂O

5X DNA loading buffer: 40% sucrose, 0.25% orange G dye, dissolved in dH₂O

Alkaline lysis solution I: 50 mM Tris-HCl (pH 8.0), 10 mM EDTA

Alkaline lysis solution II: 0.2 N NaOH, 1% (v/v) SDS,

Alkaline lysis solution III: 3 M potassium acetate (pH 5.5)

2X Kirby mix (100 mL): 2 g TPNS, 12 g sodium 4-aminosalicylate, 5 mL 2M Tris-HCl (pH 8.0), 6 mL equilibrated phenol (pH 8.0), up to 100 mL with dH₂O

Phenol:Chloroform:Isoamyl Alcohol (P:C:IAA; 50 mL): 25 mL equilibrated phenol (pH 8.0), 24 mL chloroform, 1 mL isoamyl alcohol

Protein purification

Nickel purification column :

Ni-NTA charge buffer : 50 mM NiCl₂·6H₂O

Ni-NTA binding buffer : 50 mM Tris-HCl pH 7.9, 0.5 M NaCl, 5 mM imidazole

Ni-NTA Wash buffer : 50 mM Tris-HCl pH 7.9, 0.5 M NaCl, 20 mM imidazole

Ni-NTA elution buffer : 50 mM Tris-HCl pH 7.9, 0.5 M NaCl, 1 M imidazole

Ni-NTA strip buffer: 50 mM Tris-HCl pH 7.9, 0.5 M NaCl, 100 mM EDTA

AKTA purification:

Buffer A: 5 mM imidazole, 500 mM NaCl, 20mM Tris-Cl 5 mM β-mercaptoethanol pH 7.9, protease inhibitor tablet

Buffer B: 2 M imidazole, 500mM NaCl, 20mM Tris-Cl pH 7.9, 5 mM β-mercaptoethanol

Buffer C: 10 mM Tris-Cl pH 7.9, 500 mM NaCl, 5% glycerol, 1 mM MgCl₂, 10 mM DTT, 20 μM ZnCl₂

Dialysing buffer: 10 mM Tris-HCl pH 7.9, 200mM NaCl, 0.1 mM EDTA, 1 mM MgCl₂, 20 μM ZnCl₂ and 10 mM DTT

Dilution buffer: 20 mM Tris-HCl pH 7.9, 5 mM β-mercaptoethanol, 200 mM NaCl and 5% glycerol

Gel Filtration (GF) buffer: 20 mM Tris-HCl pH 7.9, 200 mM NaCl, 5% (w/v) glycerol, 5 mM β-mercaptoethanol

Anion-exchange binding buffer: 50 mM Tris-HCl pH 7.9, 50 mM NaCl, 5% (v/v) glycerol, 5 mM β-mercaptoethanol

Anion-exchange elution buffer: 50 mM Tris-HCl pH 7.9, 1 M NaCl, 5% (v/v) glycerol, 5 mM β-mercaptoethanol

Cobalt purification column:

Equilibration/ Wash buffer: 50 mM sodium phosphate, 300 mM sodium chloride, 10 mM imidazole pH 7.4

Elution buffer: 50 mM sodium phosphate, 300 mM sodium chloride, 1 M imidazole pH 7.4

Phenylmethylsulfonyl fluoride (PMSF): 10 mg/mL in isopropanol and stored at -20°C

Western blot

1.5 M Tris-HCl pH 8.8: Tris-Base 90.855 mL pH 8.8 and 350 mL distilled water, 500 mL final volume

1 M Tris-HCl pH 6.8: Tris-Base 30.285 mL pH 6.8 and 200 mL distilled water, 250 mL final volume

Semi-dry transfer buffer (1 L): 5.82 g Tris-Base, 2.92 g glycine, 3.75 mL 10% SDS, 200 mL ethanol and make up to 1 L with distilled water

10 x TBS-Tween buffer: 12 g Tris-HCl, pH 7.6, 40 g NaCl, 5 mL Tween 20™, up to 500 mL with distilled water

Blocking solution: 5% (w/v) skimmed milk in 1 x TBS-Tween buffer, made fresh before use

Lamaelli buffer (50mL): 12.5 mL Tris-HCL pH 6.8, 4 g SDS, 5 mL 2-mercaptoethanol, 15 mL glycerol, 10 mg Bromophenol blue and make up to 50 mL with distilled water.

Running buffer: dilute (10X SDS-PAGE buffer, 1 L): 30 g Tris-Base, 144 g glycine, 10 g SDS and make up to 1 L with distilled water.

SDS-PAGE Coomassie Brilliant Blue stain: 0.25% (w/v) Coomassie Brilliant Blue (Sigma), 50% (v/v) methanol, 10% (v/v) glacial acetic acid

SDS-PAGE de-stain solution: 50% (v/v) methanol, 10% (v/v) glacial acetic acid

RNA extraction

Kirby mix (Kieser *et al.*, 2000): 1% (w/v) sodium-triisopropylphenylmethane sulphonate, 6% (w/v) sodium 4-amino salicylate, 6% (v/v) phenol made up in 50 mM Tris-HCl pH 8.3.

3 M sodium acetate, pH 6.0

5X MOPS buffer: 0.2 M MOPS pH 7.0, 0.05 M sodium acetate, 0.005 M EDTA pH 8.0 and make up to 500mL with distilled water

Sample buffer: 50% glycerol, 1 mM EDTA, 0.4% bromophenol blue

Chromatin immunoprecipitation buffers

IP buffer: 50 mM HEPES pH 7.5 with KOH, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS

IP buffer + salt: 50 mM HEPES pH 7.5 with KOH, 500 mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS

Wash buffer: 10 mM Tris pH 8.0, 250 mM LiCl, 1 mM EDTA, 0.5% Nonidet P-40 substitute, 0.5% sodium deoxycholate

5X Elution buffer: TE pH 7.6, 5% SDS

S1 protection assay buffers

S1 hybridisation buffer: 2.53 g PIPES in 90 mL distilled water, add 1.67 mL 0.5M EDTA, adjust pH 7.0 with 5 M NaOH, then autoclave. Dissolve 93.1 g NaTCA in the PIPES/EDTA solution. Make up to a final volume of 167.4 mL with sterile water, mix well and freeze at -80°C in 200 µL aliquots.

5X S1 digestion buffer: 1.4 M NaCl, 150 mM sodium acetate pH 4.4, 22.5 mM zinc acetate and autoclave, then add 100 µg/mL partially cleave non-homologous DNA (Promega DNase/RNase free herring sperm DNA) and store in 500 µL aliquots at -20°C. Make 1 X when ready to use using sterile water, chill on ice for 5 min and then add 150 units of S1 nuclease per 300 µL, mix by vortexing.

S1 stop mix: 2.5 M ammonium acetate, 0.05 M EDTA

Formamide loading buffer: 80% formamide, 10 mM NaOH, 1 mM EDTA, 0.1% xylene cyanol, 0.1% bromophenol blue.

Urea gel: 16 mL Urea gel concentrator, 29 mL Urea gel diluent, 5 mL Urea gel buffer, 400 µL 10% APS, 20 µL TEMED.

2.1.3 Strains

Strain	Genotype	Source/ reference
<i>Streptomyces coelicolor</i> M145	SCP1- SCP2-	(Hopwood et al., 1986)
<i>Streptomyces coelicolor</i> J1981	M145 $\Delta rpoC::rpoCHIS$	(Babcock et al., 1997)
<i>Streptomyces coelicolor</i> S129	J1981 $\Delta rbpA::aac(3)IV$	(Newell et al., 2006)
<i>Streptomyces coelicolor</i> J1915	M145 $\Delta glkA$ SCP1- SCP2-	(Kelemen et al., 1995)
<i>Streptomyces coelicolor</i> S115	M145 $\Delta rbpA$ Apra ^R	(Aline Tabib-Salazar)
<i>Streptomyces coelicolor</i> S401	J1915 $\Delta rbpA \Delta rbpB$ pIJ6902:: <i>rbpA</i> , pRT802:: <i>3XFLAG</i>	(Aline Tabib-Salazar)
<i>Streptomyces coelicolor</i> S403	J1915 pIJ6902:: <i>sspB</i>	(This study)

<i>Streptomyces coelicolor</i> S406	<i>J1915 ΔrbpB pIJ6902::sspB</i>	(Aline Tabib-Salazar, this study)
<i>Streptomyces coelicolor</i> S407	M145 <i>rbpA</i> ^{DAS+4}	(This study)
<i>Streptomyces coelicolor</i> S408	J1915 <i>rbpA</i> ^{DAS+4}	(This study)
<i>Streptomyces coelicolor</i> S409	M145 <i>carD</i> ^{DAS+4}	(This study)
<i>Streptomyces coelicolor</i> S410	J1915 <i>carD</i> ^{DAS+4}	(This study)
<i>Streptomyces coelicolor</i> S411	J1915 <i>rbpA</i> ^{DAS+4} / <i>carD</i> ^{DAS+4}	(This study)
<i>Streptomyces coelicolor</i> S412	<i>rbpA</i> -DAS+4, pIJ6902::sspB	(This study)
<i>Streptomyces coelicolor</i> S413	<i>carD</i> -DAS+4, pIJ6902::sspB	(This study)
<i>Streptomyces coelicolor</i> S414	J1915 <i>ΔrbpB rbpA</i> -DAS+4, pIJ6902::sspB	(This study)
<i>Streptomyces coelicolor</i> S415	J1915 <i>ΔrbpB carD</i> -DAS+4, pIJ6902::sspB	(This study)
<i>Streptomyces coelicolor</i> S416	<i>J1915 ΔrbpB rbpA</i> -DAS+4/ <i>carD</i> -DAS+4, pIJ6902::sspB	(This study)
<i>Streptomyces coelicolor</i> S417	J1981 pSET152::hrdB ^{WT}	(Chidubem Iwobi)
<i>Streptomyces coelicolor</i> S418	J1981 pSET152::hrdB ^{4xR}	(Chidubem Iwobi)
<i>Mycobacterium tuberculosis</i> M233	H37Rv <i>Tn-pip/pptr::ΔrbpA hygR</i> (Himar1 transposon integrated PI <i>pptr</i> inducible system, hyg ^R)	(Bortoluzzi et al., 2013a)
<i>Escherichia coli</i> DH5α	F-, Φ80lacZΔM15, Δ(lacZYAargF) U169, deo ^R , recA1, endA1, hsdR17(rk-, mk+), phoA, supE44, thi-1, gyrA96, relA1, λ-	Invitrogen
<i>Escherichia coli</i> ET12567 (pUZ8002)	<i>dam dcm hsdS</i> dam-13::Tn9 dcm-6 hsdM Cmr pUZ is a derivative of RK2 with a	(MacNeil et al., 1992)

	mutation in the OriT(aph)	
<i>Escherichia coli</i> ET12567 (pR9406)	dam-13::Tn9, dcm-6, hsdM, Cmr, pR is a derivative of RK2 with a mutation in the <i>oriT</i> (amp)	(Jones et al., 2013)
Rosetta II BL21	F ⁻ <i>ompT hsdS_B(r_B⁻ m_B⁻) gal dcm</i> (DE3) pLysSRARE2 (Cam ^R)	(Novagen)

Table 2.1 List of strains used throughout the study.

2.1.4 Plasmids

Vector	Details	Source/ Reference
pBlueScript II SK+	<i>E. coli</i> sub-cloning vector: <i>bla</i> , <i>lacZα</i> , Amp ^R	Stratagene Ltd. (Alting-Mees & Short, 1989)
pIJ6902	<i>S. coelicolor</i> integrative expression plasmid: <i>aac(3)IV</i> , <i>tipAp</i> , <i>oriT</i> , <i>tsr</i> , ΦC31 <i>attP</i> , Apr ^R , Thio ^R	(Huang et al., 2005; Molodtsov et al., 2018)
pRT802	<i>S. coelicolor</i> integrative plasmid: <i>aphII</i> , <i>oriT</i> , <i>lacZα</i> , ΦBT1 <i>attP</i> , Kan ^R .	(Gregory et al., 2003)
pIJ2738	Meganuclease site, I-SceI: <i>aac(3)IV</i> , <i>lacZα</i> , <i>oriT</i> , Apr ^R	(Fernández-Martínez & Bibb, 2014)
pIJ2742	Meganuclease: <i>oriT</i> , Thio ^R , <i>aac(3)IV</i> , <i>ermE</i> * <i>p</i>	(Fernández-Martínez & Bibb, 2014)
pMV306	<i>M. tuberculosis</i> integrated plasmid: Kan ^R , <i>attP</i> site, <i>oriE</i> , <i>aph</i> , ΦL5	Addgene
pSET152	Shuttle plasmid, bifunctional: replicates in <i>E. coli</i> and integrates in <i>S. coelicolor</i> , derived from pSET152; <i>aac(3)IV</i> ⁻ , <i>aadA</i> ⁺ , attP ^{ΦC31} , <i>oriT</i> _{RK2} ,	(Bierman et al., 1992)

pMT3000:: <i>rbpA</i>	pMT3000 including <i>rbpA</i> gene with its promoter region (735bp) flanked by EcoRI & KpnI	(Newell et al., 2006)
pMT3000:: <i>rbpA-NdeI</i>	pMT3000 including <i>rbpA</i> gene with its promoter region (735bp) flanked by EcoRI & KpnI <i>rbpA</i> with a NdeI site, replacing the start codon	(Aline Tabib-Salazar)
pRM1126	pET28a derivative <i>ori</i> Co1E1, <i>Kan</i> , expressing <i>M. bovis rpoC, rpoB, rpoZ, rpoA</i> with His ⁸ tag from a T7 promoter.	(Czyz et al., 2014)
pSRE3	pRT802 derivative	Sven Reisloehner (2017)
pSU904E	<i>A derivative of pRM1126, S. coelicolor rpoB, rpoC, rpoZ, rpo. ori</i> Co1E1, <i>Kan</i> ,	Simon Hodges (2016)

Table 2.2 List of plasmids used in this study.

2.1.5 Antibiotics

Name	Stock concentration	Stock made up in	Concentration used in liquid media	Concentration used in solid media
Ampicillin	100 mg/mL	60% ethanol	100 µg/mL	100 µg/mL
Carbenicillin	100 mg/mL	60% ethanol	100 µg/mL	100 µg/mL
Apramycin	50 mg/mL	dH ₂ O (Filter-sterilised)	20-50 µg/mL	20 µg/mL
Chloramphenicol	34 mg/mL	100% ethanol	34 µg/mL	25 µg/mL
IPTG	1 M	dH ₂ O (Filter-sterilised)	1 mM	0.5 mM
Kanamycin	50 mg/mL	dH ₂ O (Filter-sterilized)	50 µg/mL	25 µg/mL
Nalidixic Acid	25 mg/mL	0.15 M NaOH	25 µg/mL	25 µg/mL

Thiostrepton	50 mg/mL	Dimethyl-sulphoxide	N/A	10-15 µg/mL
X-gal	40 mg/mL	N, N-dimethyl-formamide	N/A	40 µg/mL
Diamide	0.5 M	dH ₂ O (Filter-sterilized)	0.5 mM	
Pristinamycin	3.125 mg/mL	Methanol	0.5 µg/mL	0.5 µg/mL

Table 2.3 List of antibiotics and additives used in this study including the stock/working concentrations in solid and liquid media.

2.1.6 Oligos

BL mutants in *S. coelicolor*

<i>rbpA</i> _Nde_F	5' ATGAGTGAGCGAGCTCTTCG
<i>rbpA</i> _Nde_R	5' ATGTCGTGCCTCCCGGGCTTG
<i>rbpA3KRA</i> _F	5' CGCCCGCGGCCACGCACTGGGACATGCTGATG
<i>rbpA3KRA</i> _R	5' CGGCCGCGCCTCCTCAGGGCCGTCGCCGTC
<i>rbpA3XHind</i> _F	5' CCAAGCTTCGGCTGCGGCTGCGACGACGGCGAC
<i>rbpA3XHind</i> _R	5' GAAGCTTCGCACTCTTGCGGCTGTCTCGCG

Table 2.4 List of primers designed for the BL mutations in *S. coelicolor*.

BL mutants in *M. tuberculosis*

Rv20503KA_F	5' GCCCCGCCCCGGACGCACTGGGACATG
Rv20503KA_R	5'AACGGCGGCCGGCTCGGGCAGGTCGCCCTC
Rv2050R80A_F	5' GCCACGCACTGGGACATGCTGCTGGAG

Rv2050R80A_R	5' GGGCGGCTTAACCTTCTTCGGCTCG
Rv20503KRA_F	5' GCCCCGCCCCGCCACGCACTGGGACATG
Rv20503KRA_R	5' AACGGCGGCCGGCTCGGGCAGGTCGCCCTC
Rv2050_F	5' GAATTCCTGACAACACGACCGCATC
Rv2050_R	5' AAGCTTCGCACGCGCACGACCGGACATTG
PTRF	5' GCGGTCACCTTCTACGACCTGA
RV2050R	5' GAACTCCTCGCCGTTGTCGGTG
MTB_F2	5' CGACCGAGCGCAACGCGTGCGGCCGC
MTB_R2	5' GATCGTACGCTAGTAACTACGTCGACA

Table 2.5 List of primers designed for the BL mutations in *M. tuberculosis*.

Multiplex sequencing Illumina RNA-seq adapter oligos

Illumina multiplexing adapter 1	GATCGGAAGAGCACACGTCT
Illumina multiplexing adapter 2	ACACTCTTTCCCTACACGACGCTCTTCCGATCT

Table 2.6 List of multiplex sequencing Illumina RNA-seq adapter oligos

S1 protection assay

S1_promoter_oligo	5' GCACCACGCCGGTGAACAGCTCTTCGCCTTTGCTCACCATATCGAAGCTTTGTCGGTG
SpeI-ypet-for	5' GAACTAGTACAGCTATGACATGATTACGAA
EcoRV-ypet-Rev	5' GCGATATCTCACCTCGAAAGCAAGCTGAT

Table 2.7 Primer designed for S1 nuclease protection assay

In vivo expression of ribosomal promoters in DAS depletion strains

rrnDp2_nt	5' GATCCTTCCC GCAAGAGCCGTTGACACGGAGCGAGCGGGGAGGTA
rrnDp2_t2	5' AGCTTACTGTT CGAATCTACCTCCCCGCTCGCTCCGTGTCAACGG
rrnDp3_nt	5' GATCCCGGAAACGAAGGCCGGTAAGACCGGCTCGAAAGTTCTGAT
rrnDp3_t2	5' AGCTTGCGGCTCCGACTTTATCAGAACTTTGAGCCGGTCTTACC
rrnDp2-3ch_nt	5' GATCCTTCCC GCAAGAGCCGTTGACACGGCTCGAAAGTTCTGATA
rrnDp2-3ch_t2	5' AGCTTACTGTT CGAATCTACCTCCCCGCTCGCTCCGGTCTTACGG
rrnDp3-2ch_nt	5' GATCCCGGAAACGAAGGCCGGTAAGACCGGAGCGAGCGGGGAGGT
rrnDp3-2ch_t2	5' AGCTTGCGGCTCCGACTTTATCAGAACTTTGAGCCGTGTCAACG

Table 2.8 Primers designed for the in vivo study of the ribosomal promoters in the RbpA/CarD-DAS depletion system.

σ^{HrdB} association with RNAP qPCR targets

<i>hrdB</i> -double_F	5' GAGGCCGGTGTCGCGCCGAGGACAAGCTGGCGAACTC
<i>hrdB</i> -double_R	5' GATGCGGGTGCGGAGGCGGACCTCCTGCTCGGCGTTGAG
3SCO5281_F	5' GTAGGCGTAGCCTTGATTCC
3SCO5281_R	5' CTCGAGTTACTGGGGGACTG
8SCO5281_F	5' CGACCTGATCGACAAGAAGC
8SCO5281_R	5' GCCTCTCCAGGGTGATGT

Table 2.9 Primers designed for σ^{HrdB} association with RNAP using qPCR

ChIP-seq adapter

ChIP Illumina adapter	5' AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
-----------------------	--

Table 2.10 Adapter using for ChIP-seq library preparation

DAS depletion

<i>rbpA</i> UF Extension	5' GTCTACGGTTCCGAGGAACTG
<i>rbpA</i> INT + Spe1	5' ACTAGTCCCTGTCCCGCTCCCTGTTC
<i>rbpA</i> UR	5' GCGTAGTTCTCGGAGTAGTTCTCGTCGTTGGCGGCCGCACTCTTGCGGCTGTCTC
<i>rbpA</i> DF	5' GACGAGAACTACTCCGCGAACTACGCCGACGCCTCCTAGCGGGAAACCGTAGGCGAC
<i>rbpA</i> DR INT +NOT1	5' AGCGGCCGCGAGGTGCCGATCACGTTCTGTG
<i>rbpA</i> DR Extension	5' GGTGGTCAACTGGCCCAAGTC
<i>carD</i> UF extension	5' CACGGCGGCGGCAACTCCTTG
<i>carD</i> UF INT+ SPE1	5' ACTAGTCTCGTCCGACATTCAGCTATG
<i>carD</i> UR	5' GAGCGTAGTTCTCGGAGTAGTTCTCGTCGTTGGCGGCGGAGGCGAGCACCTCGTC
<i>carD</i> DF	5' GACGAGAACTACTCCGAGAACTACGCCGACGCCTCCTGATCCGACTGGATCCGGGGCGA
<i>carD</i> DR INT + NOT1	5' AGCGGCCGCGTGACGTCGTCGGTCACCGTC
<i>carD</i> DR EXTENSION	5' GCTCACGCCCGTCCTCGAAGGC
<i>carD</i> -DAS_F	5' CGTCGTTACAAGGCAAATCTG
<i>carD</i> -DAS_R	5' GTGCGCTGAACCGAGTGGGTG
<i>rbpA</i> -DAS_F	5' CGATTCGAGATGCCCTTCTCG
<i>rbpA</i> -DAS_R	5' GTTCGGTTGTGTCGCCTACGG

pRT802: <i>rbpA</i> F	5' GAAATCGTTAGTTAGGCTAAGGAGTCGGAACGGGAATCTTTACCGCC
<i>carD</i> - <i>rbpA</i> R	5' TTCGCAGGTCCTGCGGGCCGGGCACGGC
<i>rbpA</i> - <i>carD</i> F	5' CGGCCCCGAGGACCTGCGAATACTTTCTTCC
<i>carD</i> -pRT802R	5' CTATGGTGACGAAGGAACTAGCTGAACCGAGTGGGTGA

Table 2.11 List of the primers designed for the DAS dependant depletion system.

FLAG-DAS strains

R/C_FLAGDASinvF	5' GACATCGACTACAAGGACGATGACGATAAGGGATCCGCCCAACGACGAGAAC
<i>rbpA</i> FLAGDASinvR	5' ATGGTCTTTGTAGTCGCCGTCGTGGTCCTTGTAGTCAAGCGCACTTTGCGGCTGTC
<i>carD</i> FLAGDASinvR	5' ATGGTCTTTGTAGTCGCCGTCGTGGTCCTTGTAGTCAAGGGAGGCGAGCACCTCGTC

Table 2.12 Primers designed for the FLAG-DAS strains.

Application of the DAS depletion system *in vivo*

1-pRT802: <i>rbpA</i> F	5' GAAATCGTTAGTTAGGCTAAGGGAGTCGGAACGGGAATCTTTACCGCC
2- <i>carD</i> - <i>rbpA</i> R	5' TTCGCAGGTCCTGCGGGCCGGGCACGGC
3- <i>rbpA</i> - <i>carD</i> F	5' CGGCCCCGAGGACCTGCGAATACTTTCTTCC
4- <i>carD</i> -pRT802 R	5' CTATGGTGACGAAGGAACTAGCTGAACCGAGTGGGTGA

Table 2.13 Primers designed for validation of the DAS system

Recombinant RNA polymerase

NSB_Link_F	5' GGCCGCCACACACTAGTCACCAACCG
NSB_Link_R	5' GATCCGGTTGGTGAAGTAGTGTGTGTGGC

RpoA_F	5' CAATCCCCCTCTAGAGTTTAAACTTTAAGAAGGAGATATACTCATGCTGATCGCTCAGCGTC
RpoA_R	5' CTTATATTGTGTTAATTGCTTAATTAAGTTAGTACTGCTCGGTCTCCACGA
RpoZ_F	5' CAATTAACACAATATAAGAGGAGGTCCAGTCGTGTCCTCTTCCATCTCCGCG
RpoZ_R	5' CTAGTGTGTGTGGCGGCCGCTTACTGCGCCGGGCCCTCGATG
Rpo_C1_Fwd	5' CTGGCGCGCCAGACTAGTGGCTCGGGTGCAATGCTCGACGTCAACTTCTTCG
Rpo_C3_Rev	5' ACCGCATGCATGGACGACAACCTGGGATCCCTAGTGATGGTGATGGTGATGGTGATGCTGGTTGTAC GGACCGTAG
RpoB1_fwd	5' CCATCGAGGGCCCGGCGCAGTAAGCGGCCGCTTTAAGAAGGAGATATATCTATGGCCGCCTCGCGCA ATGC
RpoB3_rev	5' ACCGCATGCATGGACGACAACCTGGGATCCCGAGCCACTAGTCTGGCGCGCCAGGACCTCTTCGACGC TGCTC
Rpo_C1_rev	5' TGCGGCCGCGCTCGACCATGCGCTTGCGCT
Rpo_C3_rev	5' ACCGCATGCATGGACGACAACCTGGGATCCCTAGTGATGGTGATGGTGATGGTGATGCTGGTTGTAC GGACCGTAG

Table 2.14 Primers designed for *S. coelicolor* recombinant RNAP expression.

2.2 Growth selection and storage of bacterial strains

2.2.1 Liquid and solid media

L agar (LA): 10 g Difco Bacto tryptone, 5 g Oxoid yeast extract, 5 g NaCl, 1 g glucose, 10 g agar made up to 1 L with distilled water.

Lennox broth (LB): 10 g Difco Bacto tryptone, 5 g Oxoid yeast extract, 5 g NaCl, 1 g glucose, made up to 1 L with distilled water.

Mannitol soya flour (MS) agar: 20 g agar, 20 g mannitol, 20 g soya flour, made up to 1 L with tap water.

YEME liquid medium: 3 g yeast extract, 5 g Difco Bacto peptone, 3 g malt extract, 10 g glucose or glycerol, 100 g sucrose, made up to 1 L with distilled water.

2 X PG medium; per 100 mL: 1 g Difco yeast extract and 1 g Difco Casamino acids. 5 mL aliquots were autoclaved.

2x YT Broth: 1.6 g Bacto Tryptone, 1 g Bacto Yeast extract, 0.5 g NaCl, pH adjusted to 7.0 using 5 N NaOH, made up to 90 mL with distilled water.

Difco nutrient agar (DNA): 4.5 g Difco nutrient agar, distilled water 200 mL, 4.5 g agar.

Supplemented liquid minimal medium, liquid (SMM): 81.9 mL PEG 6000, 2.5 mL (5mM final) $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 10 mL (25 mM final) TES buffer (0.25 M, pH 7.2), 1 mL (1 mM final) $\text{NaH}_2\text{PO}_4 + \text{K}_2\text{HPO}_4$, 2 mL glucose (50% w/v), 1 mL Antifoam 289 (1% w/v), 0.1 mL Trace element solution, 1 mL Difco casaminoacids (20% w/v), make up to 100 mL with sterile distilled water.

Supplemented minimal medium, solid (SMMS): A minimal, solid medium used for production of *S. coelicolor* antibiotics. 2 g Difco casamino acids, 5.73 g TES buffer, dissolved in 1 L dH₂O and adjusted to pH 7.2 with 5 N NaOH. Typically divided into 200 mL aliquots in 250 mL Erlenmeyer flasks with 3 g agar, stoppered with a foam bung and foil, and autoclaved. Prior to use add 2 mL 50 mM $\text{NaH}_2\text{PO}_4 + \text{K}_2\text{HPO}_4$ (1 mM final), 1 mL 1M MgSO_4 (5mM final), 3.6 mL 50% (w/v) glucose (50 mM final) and 0.2 mL SMM trace element solution (see below).

SMM trace element solution: 0.1 g/L each of $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$, $\text{CaCl}_2 \cdot 6\text{H}_2\text{O}$, and NaCl, made fresh over 2-4 weeks, stored at 4 °C.

Minimal liquid medium (NMMP): Media used to obtain dispersed cultures with specific carbon sources. 2 g Ammonium sulphate, 5 g Difco casaminoacids, 0.6 g Magnesium sulphate, 50 g PEG 6000, 1 mL minor elements and 800mL distilled water.

Minor elements: 1 g Zinc sulphate, 1 g Iron sulphate, 1 g Manganese chloride, 1 g Calcium chloride.

Minimal media: 0.5 g L-asparagine, 0.5 g K_2HPO_4 , 0.2 g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.01 g $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ and make up to 1 L with distilled water, separate into 200 mL aliquots and add 3 g Difco Bacto agar to each flask pH 7-7.2 and autoclave, then add 4 mL 50% glucose or glycerol.

5X M63 minimal medium: 5 g $(\text{NH}_4)_2\text{SO}_4$, 34 g KH_2PO_4 , 1.25 mg $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 2.5 mg vitamin B1 (thiamine) and make up to 500 mL with distilled water, adjust pH to 7.0 with KOH.

RG-2 media pH 7: 50 mM glycerol, 200 mM sucrose, 5mM K_2HPO_4 , 80 mM KCl, 4 mM MgSO_4 , 50 mM TES (KOH), 10% antifoam and 0.75 mL/ of RG-2 media.

ADC: 10 g Bovine albumin fraction V, 4 g Dextrose, 0.008 g Catalase, 1.7 g sodium chloride, make this supplement up to 200 mL with distilled water and filter sterilise.

OADC: 10 g Bovine albumin fraction V, 4 g Dextrose, 0.008 g Catalase, 1.7 g sodium chloride, 0.1 g oleic acid, make this supplement up to 200 mL with distilled water and filter sterilise.

7H9: 1.88 g 7H9, 360 distilled water and 200 μL Tween 80, autoclave and add 40 mL ADC (total volume 400 mL).

7H10: 7.6 g 7H10, 2 mL glycerol, 360 mL distilled water and autoclave, cool to 45°C and add 40 mL OADC (total volume 400 mL).

2.2.2 Growth and storage

E. coli

General growth conditions: *E. coli* were grown on solid or liquid media at 37 °C for up to 24 h. For liquid culture, *E. coli* were grown in LB (with selection) in an orbital shaker at 250 rpm. For solid culture, *E. coli* were grown on LA (with selection) and once colonies formed plates were stored at 4 °C for up for 2 weeks.

Making glycerol stocks: For long term storage of *E. coli* strains were stored in a 20% glycerol solution and stored at -80 °C.

Streptomyces

General growth conditions: *Streptomyces* was grown on solid media at 30 °C for 5-7 days until colony formation and sporulation was observed.

Growing *S. coelicolor* liquid cultures: Spore stock concentrations were measured using optical density and the required volume to give a starting OD_{450} of 0.05 was added 1 mL 0.05M pH 8

TES buffer. The mixture was briefly centrifuged, the supernatant was removed, and the spore pellet was re-suspended in 1 mL of TES buffer. The spore suspension was heat shocked at 50 °C for 10 min, mixed with 1 mL 2 X PG media and 2 µL 5 M CaCl₂, and incubated in a universal tube at 37 °C with vigorous shaking for 3h. Following pre-germination of spores, the culture was briefly centrifuged and re-suspended in 1 mL liquid media. The suspension was used to inoculate the *S. coelicolor* liquid culture which was cultured at 30 °C in an orbital shaker at 300 rpm. *S. coelicolor* were grown in siliconized Erlenmeyer flasks with a stainless-steel spring loop in the base of the flask to prevent formation of mycelium clumps and improved aeration.

Making spore stocks: A single colony was streaked to confluency on MS agar and incubated at 30 °C for 5 days until the mycelium turned dark grey, indicating production of mature spores. To harvest spores, 9 mL sterile water was added to the plate and scraped with a metal loop suspending spores in the water. The crude suspension was filtered through a "spore tube" (2 cm of non-absorbent cotton wool in a 10 mL syringe, autoclaved to sterilise) into a 15 mL tube and centrifuged at 1,000 x g for 10 min. The supernatant was discarded, and the pellet was re-suspended in 1-2 mL sterile 20% glycerol. Spore stocks were stored at -20 °C and remain viable for years, even if repeatedly thawed and frozen.

2.3 DNA manipulation

2.3.1 PCR

Each reaction consisted of 10 µL 5X Phusion HF (New England Biolabs), 1 µL 10 mM dNTPs, 2.5 µL 10 µM of each primer (Eurofins), 100 ng genomic M145 DNA a prototrophic derivative of strain A3 (2), 1.5 µL DMSO (100X; Fisher Scientific), 0.5 µL of Phusion DNA polymerase (New England Biolabs). The final volume was made up to 50 µL by adding nuclease free water (autoclaved) ultra-pure water.

The thermocycler conditions:

Initial denaturation: 98° C for 30 sec

Denaturation: 98 °C for 5-10 sec

Annealing: Lowest Primer T_m °C for 10 -30 sec

Extension: 72°C for 15 sec/kb

Final extension: 72°C for 5-10 min

Hold at 4°C

The thermocycler conditions for the annealing and extension were altered for the optimisation of the primers used here.

Inverse PCR

Primers were phosphorylated by combining 200 pmol oligonucleotides, 1 x T4 DNA kinase buffer, 2 mM ATP and 0.005 units T4 DNA kinase. The reaction was left at 37°C for 30 min and then heated at 90°C for 2 min to inactivate the enzyme. The primers were used to set up a PCR reaction as stated in Section 2.2.1.1, using Accuzyme (Bioline) as the thermostable DNA polymerase.

For large products (>3kb), the following PCR cycle was used:

- Melting at 96°C for 1 min
 - Annealing at 55-65°C for 45 s
 - Extension at 72°C for 2.5 min per 1 kb
 - Melting at 96°C for 1 min
 - Annealing at 55-65°C for 45 s
 - Extension at 72°C for 3.5 min per 1 kb
 - Final Extension at 72°C for 15 min
 - Held at 4°C
- } x 10
- } x 10

The PCR products were then separated on a 0.8 to 1% of agarose gel and the gene of interest was purified using QIAquick™ PCR Purification Kit (Qiagen) as per manufacturer's protocol.

Two-stage PCR

PCR primers were designed to amplify the upstream and downstream flanks separately, with an integrated DAS tag, after the gene sequence (stop codon removed). Once these flanks with the DAS tag were amplified separately, these were diluted (1/10), and used as a template for the second stage of the PCR. For the second stage, primers were designed internal to the external primers from the first stage, with restriction sites attached. The internal primers should amplify 1000 bp upstream of the gene + the gene + the DAS tag + 1000 bp downstream of the gene.

Colony PCR

M. tuberculosis

The strains were cultured in Middlebrook 7H9 medium supplemented with ADC, Tween 80, kanamycin, hygromycin and pristinamycin for 7 days. 1mL of this culture was heat killed by boiling for 1 h and a FastDNA spin kit for soil was used following the manufacturer's instructions for genomic extractions. The genomic DNA was used for PCR and the resulting PCR products were sequenced.

Alternatively, a Chloroform method was implemented, firstly, *M. tuberculosis* was heat killed by boiling for one h at 99°C. Equal volumes of chloroform was added, vortex and centrifuge for one minute at full speed. The aqueous layer was harvested which contains the genomic DNA.

S. coelicolor

A single colony was picked and re-suspend in 0.5% SDS, the mixture was vortexed vigorously and heated to 98°C for 5 min. The mixture was centrifuged, and the supernatant was collected into a clean 1.5 mL microcentrifuge tube. To use this genomic preparation, 1:5 dilutions were prepared and add to a PCR reaction.

2.3.2 Gibson Assembly

For 2-3 fragments, 10 µL of Gibson Assembly Master mix (2X; New England Biolabs), 1µL of the fragments were added and the final volume was made up to 20 µL. The samples were

incubated in a thermocycler at 50°C for 2 h and store at -20°C. This mix was then be transformed into *E. coli* DH5α cells.

2.3.3 Restriction digest

All digests were done in a final volume of 10-40 µL, this includes the appropriate amount of DNA, 0.5-1 µL of the restriction enzyme (20,000 units/mL; NEB), restriction buffer (diluted from a 10x stock), the final volume was made up with MQ water. The digests were incubated for a minimum of 3 h. The digests were confirmed by gel electrophoresis, by running undigested DNA.

2.3.4 Gel electrophoresis

0.8% or 1% agarose (Sigma) gels were used to analyse undigested and digested plasmids. The gel was prepared using 100 mL 1X TBE buffer to separate the 1.5 kB flanking regions. The samples were run at 150 mA and 75 V against a 1 Kb hyperladder (Bioline). To visualise the DNA, gels were stained with GelRed™ Nucleic Acid Gel Stain (Biotium) according to the manufacturer's instructions. The gel was placed onto a rotator (IKA® KS 260 Basic) for 15 min, followed by visualisation using a high UV transilluminator.

The same method from above was used to visualise DNA fragments, however the samples were visualised under the low UV transilluminator to limit the mutagenic properties of UV light. The desired band was excised and purified using the Gel Purification kit (Qiagen), following the manufacturer's instructions.

2.3.5 Gel purification

To purify DNA fragments following gel electrophoresis, agarose gel fragments were excised with a scalpel and purified with the Wizard® SV Gel and PCR Clean-Up System (Promega) according to manufacturer's instructions.

2.3.6 DNA dephosphorylation

DNA was dephosphorylated in 1 X Antarctic Phosphatase buffer with 5 units Antarctic Phosphatase (New England Biolabs) and incubated at 37 °C for 15 min (60 min for 3' extensions) before heat inactivation at 70 °C for 5 min.

2.3.7 DNA ligation

The ligation reaction consisted of 1 µL 10x T4 DNA ligase buffer and 1 µL T4 DNA ligase, approximately 100 ng of vector DNA with an appropriate concentration of insert, see below:

$$\text{Insert (ng)} = (100 \text{ ng} \times \text{size of insert (kb)}) / (\text{size of vector (kb)}) \times 3/1$$

2.3.9.1 Small scale plasmid isolation from *E. coli*

To purify plasmid DNA from *E. coli*, cells were harvested from 3 mL overnight culture at 16,000 x g for 5 min and the cell pellet was resuspended in 200 µL alkaline lysis solution I. 400 µL alkaline lysis solution II was added, mixed by inverting 5 times before addition of 300 µL alkaline lysis solution III + 1 µL 10 mg/mL RNase A and incubation for 10 min at room temperature. The mixture was centrifuged at 16,000 x g for 10 min and the supernatant was transferred to a new tube. 150 µL P:C:IAA was added, vortexed for 2 min, centrifuged at 16,000 x g for 10 min and the upper phase was transferred to a new tube. 600 µL isopropanol was added, chilled on ice for 10 min and centrifuged at 16,000 x g for 10 min. The supernatant was removed, and the pellet was washed with 200 µL 70% ethanol before centrifugation at 16,000 x g for 1 minute. All supernatant was removed, and the pellet was resuspended in 50 µL dH₂O.

2.3.9.2 Large Scale plasmid isolation from *E. coli*

To purify plasmid DNA from *E. coli* on a large scale, the QIAGEN Plasmid Midi Kit (Qiagen) was used according to manufacturer's instructions.

2.3.9.3 Transformation of chemically competent *E. coli* DH5α

The transformation protocol (Invitrogen) was used to insert the ligation mix into the competent *E. coli* DH5α cells. 5 µL of the ligation mix was added to a 50 µL aliquot of the competent *E. coli* DH5α cells. This was left to stand on ice for 30 min, followed by heat shocking at 42 °C for 20 sec without shaking. The sample was placed on ice for a further 2 min and 950 µL of pre-warmed LB was added to the microcentrifuge tube. This was incubated at 37 °C for 1 h at 225 rpm. A 100 µL sample was spread over pre-warmed LA agar plates (with appropriate antibiotics) and incubated overnight at 37 °C. Typically, transformants were seen 15-17 h post overnight incubation.

2.3.9.4 Preparation of chemically competent *E. coli* (CaCl₂ method)

The preparation of competent *E. coli*, DH5α cells was performed by following the manufacturers protocol (Invitrogen). A single colony was transferred to 5 mL of LB broth with appropriate antibiotics and grown for overnight at 37 °C shaking. 1:100 of the overnight culture was used to inoculate a fresh LB broth flask the following morning and the cells were left to culture to an OD600 0.4-0.5. The cells were harvested by centrifugation at 2,000 x g for 10 min and the pellet was re-suspended in 20 mL ice-cold 0.1 M MgCl₂ 10% glycerol. The cell suspension was centrifuged at 2,000 x g for 10 min and the pellet re-suspended in 10 mL ice-cold 0.1 M CaCl₂ 10% glycerol. The solution was incubated on ice for 30 min before being centrifuged for a further 10 min at 2,000 x g. The cell pellet was re-suspended in the residual volume, approximately 3 mL and 150 µL aliquots were frozen in dry ice and stored at -80 °C.

2.3.9.5 Intergeneric conjugation of DNA into *S. coelicolor*

Intergeneric conjugation into *S. coelicolor* was carried out using *E. coli* strains which contain a vector with the *oriT*, ET12567/ pUZ8002 and ET12567/pR9407. A single colony was used to inoculate a 5 mL LB culture with the appropriate antibiotics overnight at 37 °C. This culture was then used to inoculate a 15 mL culture and grown to an OD600 of 0.4-0.6. The cells were then washed twice with cheap muller broth and re-suspended with 0.1 volume of LB broth. In the meantime, 5 µL of the appropriate *S. coelicolor* strain was added to 500 µL of YT broth and

heat shocked at 50°C for 10 min. Once cool, 500 µL of the *E. coli* cells were added to the YT broth mixture. The cells were spun, followed by re-suspending the pellet in the residual volume (~100 µL). This was then plated onto pre-dried (for one h) MS agar + 10 mM MgCl₂ plates. After 16-20 h an overlay was poured over the plate which contained 1 mL of sterile water and Naladixic acid with other antibiotics for the selection of the plasmids and incubated at 30°C. Following 3-4 days, exconjugants were picked off for further plating.

2.3.9.6 Electroporation of *M. tuberculosis*

10 mL 7H9-Tween was inoculated with the *M. tuberculosis* RbpA BL strains were grown to log phase. 2 M glycine was added to a final volume of 1.5% v/v and incubated for 16-24 h. The cultures were centrifuged for 10 min and the pellet was washed twice with 10 mL 10% glycerol followed by 5 mL 10% glycerol. The cell pellet was re-suspended in 1 mL 10% glycerol and 2 µg of plasmid DNA was added to 0.2 mL of the *M. tuberculosis* culture. The culture and DNA were mixed thoroughly and transferred to a 0.2 cm electrode gap electroporation cuvette. Cells were electroporated with a single pulse of 2.5 kV, 25 µF and 1000 Ω. After the electroporation, cells were transferred immediately to 7H9 broth and incubated for a further 16 hours at 37°C. The cells were harvested by centrifugation for 3,000 x g for 10 min and the cells were diluted to 1:10 and 1:100, 100 µL of these dilutions were plated onto 7H10-OADC supplemented with 30 µg/mL kanamycin, 50 µg/mL hygromycin and 0.5 µg/mL PI. The plates were dried and incubated for 4 weeks at 37°C.

2.4 Setting up cultures

M. tuberculosis

The basic linker mutants were cultured in the presence of 0.5 µg/mL Pristinamycin (PI) for 7 days and harvested at 3,000 rpm for 10 min. The supernatant was removed, and the cells were washed twice in PBS (to remove pristinamycin) followed by normalising the amount of cells added to a fresh 10 mL Middlebrook 7H9 media (without PI) to ensure that the starting point for each culture was similar. The optical density (OD₆₀₀) was measured at regular intervals and colony forming units (CFUs) were measured by plating a dilution series of bacilli onto

Middlebrook 7H10 agar supplemented with OADC, 30 µg/mL kanamycin and 50 µg/mL hygromycin. The cultures were plated onto LB agar on day 28 to check for contamination, no growth was seen on any plates after overnight incubation. Colony PCR was repeated at regular intervals through the experiments to check that use of PI had not caused spontaneous PI resistance, therefore on each occasion the *ptr* promoter and Rv2050 downstream of the transposon insertion were amplified and sequenced.

S. coelicolor

Normalise the spore stocks to OD₄₅₀ 0.05 for all of the spores to ensure that the starting number of spores between samples is the same. The spores are added to 1 mL 0.05 M TES buffer (pH check), and heat shocked for 10 min at 50°C. The heat shocked spores are added to 1 mL 2 X PG media and 2 µL 5 M CaCl₂ and incubated for 3 h at 37°C shaking. The spores are vortexed vigorously and added to a desired amount of media (YEME, NMMP, SMM) and left overnight.

2.5 Most probable number assay (MPN)

MPN is used to estimate the minimal number of cells required for growth, by serially diluting the bacterial culture. Firstly, *M. tuberculosis* was cultured to mid log phase (day 7, 0.2 OD₆₀₀) in the presence of PI (0.5 µg/ml), kanamycin (30 µg/ml) and hygromycin (50 µg/ml). Secondly, 450 µL 7H9 media supplemented with kanamycin and hygromycin was pipetted into every 24-well plate. Thirdly, 50 µL samples from each RbpA BL strain was placed into the first well (10⁻¹) and serially diluted to 10⁻¹², this was carried out as four technical replicates and three biological replicates for M233 (pMV306::*rbpA*^{WT}), M233 (pMV306), M233 (pMV306::*rbpA*^{3KA}), M233 (pMV306::*rbpA*^{R80A}) and M233 (pMV306::*rbpA*^{3KRA}). The plates were incubated for six weeks at 37°C and then read using a plate reader (OD₆₀₀). A control plate with no cells was used as a measure of no growth. The 12 readings per strain were entered into an MPN calculator to generate an MPN score

2.6 Colony forming units (CFU) for *M. tuberculosis*

A serial dilution method similar to the MPN assay was used to determine the colony forming units per mL (CFU/mL) of *M. tuberculosis*. 100 µL of the liquid culture was added to 900 µL of sterile PBS (10^{-1} dilution), vortexed briefly. 100 µL from the 10^{-1} dilution was further diluted in 900 µL PBS and so on until a dilution series from 10^{-1} - 10^{-12} was generated. 40 µL samples from each dilution series was spotted onto 7H10 agar plates supplemented with kanamycin and hygromycin. These plates were incubated for four weeks at 37°C and then the number of colonies/mL were calculated using the following equation:

$$\frac{CFU}{mL} = \frac{\text{no. colonies} \times \text{dilution factor}}{\text{Vol plated in mL}}$$

2.7 Protein purification

Ni-NTA spin column: Ni-NTA spin column (Qiagen) was used for small-scale purification of His-tagged proteins according to manufacturer's instructions.

TCA precipitation: Mycelium (1.5 mL) were sampled at 3-time points (0, 20 and 40 min), upon induction three different stress (Diamide 1 mM, serine hydroxymate 50 mM and sodium hypochlorite 2.5 mM), centrifuged at 16,100 x g for 10 min at 4 °C. The pellets were then washed with 240 µL ice-cold 0.9% (w/v) NaCl followed by 60 µL of 50% TCA and incubated on ice for 20 min. This was then centrifuged again at 16,100 x g for 10 min at 4°C, the TCA was removed, and microcentrifuge tubes were spun again as above. The pellets were washed in acetone to ensure that all the TCA was removed and spun again as above. The acetone was left to evaporate followed by re-suspending the pellet in 100 µL of loading buffer (100 mM iodoacetamide, 6 M urea, 200 mM Tris/HCL, pH 7.2, 10 mM EDTA and 0.5% SDS). Samples were boiled and loaded on a 12% Bis-Tris non-reducing protein gel.

Cobalt Histag columns: followed manufacturers protocol (Thermo Scientific)

HisTrap purification: The HisTrap nickel column was washed with 5 CV distilled water, 5 CV buffer A (refer to section 2.2.1 Methods), equilibrate with 5 CV buffer B (Methods section 2.2.1) and wash with 10 CV buffer A until the UV detector is zero. The sample was loaded onto

a 5 mL loading loop using the injection function and loaded, several times until all of the cell lysate was injected onto the column. The column was washed with 5 CV buffer A, then the protein was eluted into 0.5 mL fractions using the fractionator with an imidazole gradient was completed. 10 μ L of these fractions were run on an western gel and the fractions confirming the presence of the protein were pooled together and concentrated using a ultracentrifugation vivaspin column (Sartorius). The protein sample was dialysed in buffer C (Methods 2.2.1) overnight.

2.8.1 Gel filtration

Gel filtration was performed using a 200 μ g 16/600 column (GE Healthcare) and Akta FPLC (Amersham PLC). The column was washed with 10 CV of gel filtration buffer overnight, pump washed and 2mL of concentrated protein sample was loaded, the program S16/600 fractionation was run. 10 μ L samples for each of the fractions was run on an SDS-PAGE gel and the fractions confirming presence of the protein were pooled together and concentrated to 250 μ L and re-dialysed using buffer C overnight at 4°C.

2.8.2 Protein sample analyses via SDS-PAGE

Every protein sample was normalised to 10 μ g, 10% SDS gels Tris-glycine SDS- polyacrylamide gel electrophoresis was used to run the protein samples. Protein samples were mixed with the appropriate volumes of DTT and loading buffer. The protein samples were boiled at 98°C for 10 min and centrifuged at full speed for 1 min. Gels were run using a Hoefer SE260, a mini vertical gel electrophoresis unit (Amersham Biosciences) at 120 V for 2 h in NuPAGE® MES SDS running buffer (Invitrogen). Protein was visualised by staining with Ponceau stain, followed by Coomassie Brilliant Blue and washed with de-staining solution.

2.8.3 Western blot analysis

The protein samples were separated by SDS-PAGE and transferred onto a Protran® nitrocellulose membrane (0.2 μ M, Whatman), using a semi-dry transfer method at 20 V for 1 h. Following this, the membrane was washed with 1x TBS-Tween 3 times for 10 min and

incubated with blocking solution at RT for 1-2 h with gentle rocking at 4°C. The membrane was washed again 3 x 10 min with 1x TBS-Tween and incubated overnight with primary antibody, which was diluted in blocking solution at 4°C with gentle rocking. Following this, the membrane was washed for 3 x 10 min with 1 x TBS-Tween and incubated for 1 h with the secondary antibody HRP conjugate which was also diluted in blocking solution at 4°C with gentle rocking. The membrane was washed for 3 x 10 min with 1 x TBS-Tween and subjected to ECL detection. The ECL developer solutions A and B were mixed in a 50:50 ratio and poured onto the membrane for 1 min, the membrane is then imaged using the LICOR.

2.9 RNA isolation from *S. coelicolor*

The *S. coelicolor* cultures were grown to OD₄₅₀ 0.6, at which point a stop solution (95:5 ethanol: acid phenol) was added to the cultures in a 5:1 ratio (culture: stop solution). 10 mL cultures were left at RT for 10 min and centrifuged to remove the supernatant. 2 volumes of RNA protect was added to the samples, vortexed for 5 sec and incubated at RT for 5 min. The samples were centrifuged for 10 min at 5000 x g and the supernatant was removed. 1 mL of TRI reagent was used to resuspend the pellet, before flash freezing. Once all of the samples were harvested, the cells in 1 mL TRI reagent were thawed and the nucleic acid extraction protocol was followed. 200 µL TRI reagent was added to all of the samples and placed into 0.1 mm silica beads (Lysing matrix B) and mixed briefly. The bacterial cells were lysed using the ribolyser for 45 sec at 6.5 Hz. The samples were left at RT for 10 min and then 200 µL chloroform was then added to each cell suspension, vortexed for 30 sec, and incubated at RT for a further 10 min to partition the aqueous and phenolic phases before centrifuging at 13,000 rpm for 15 min at 4°C. The aqueous phase was then transferred to a fresh microcentrifuge tube and re-extracted with an equal volume of chloroform. The aqueous phase was transferred to a fresh microcentrifuge tube, 0.8 volume isopropanol added and mixed by inverting twice and incubated overnight at -20°C to precipitate the nucleic acids. The nucleic acid extractions were centrifuged at 13,000 rpm for 20 min at 4°C. A white precipitate should be visible, the supernatants were carefully removed and discarded. The pellets were washed by adding 500 µL cold 70% ethanol, centrifuging at 13,000 rpm for 15 min at 4°C and removing the ethanol, the tubes were re-spun to remove any excess ethanol. The pellets were air dried for a few min

and re-suspended into 100 µL RNase free- water. The RNA was cleaned up using a RNeasy column (following manufacturer's instructions). The RNA was stored at -70°C and a 5 µL sample was stored separately for bioanalysis of the RNA quality.

2.9.1 RNA isolation from *M. tuberculosis*

M. tuberculosis was cultured in 50 mL Middlebrook 7H9 supplemented with ADC, Tween 80, 30 µg/mL kanamycin and 50 µg/mL hygromycin for 8 days and the OD₆₀₀ was measured to ensure that all of the biological replicates were in the same growth phase. Transcription was stopped in the cultures by adding 4 volumes of 5 M GTC (5 M guanidine thiocyanate, 0.5% sodium N-lauroyl sarcosine, 25 mM sodium citrate, 1% Tween-80, 0.1 M β-mercaptoethanol; Salina et al., 2014). The cultures were centrifuged in 30 mL aliquots for 30 min and the supernatant was carefully removed. The process was repeated until the bacilli were combined and re-suspended into 1 mL of Trizol and stored at -80°C. The sterilised samples were removed from the CL3 laboratory and each of the samples were made up to 1.2 mL Trizol and added to 0.5 mL of 0.1 mm silica beads (Lysing matrix B). The bacterial cells were lysed using a ribolyser for 45 sec at a speed of 6.5 Hz. The disrupted cells were then left at room temperature for 10 min. 200 µL chloroform was then added to each cell suspension, vortexed for 30 sec, and incubated at RT for a further 10 min to partition the aqueous and phenolic phases before centrifuging at 13,000 rpm for 15 min at 4°C. The aqueous phase was then transferred to a fresh microcentrifuge tube and re-extracted with an equal volume of chloroform. The aqueous phase was transferred to a fresh microcentrifuge tube, 0.8 volume isopropanol added and mixed by inverting twice and incubated overnight at -20°C to precipitate the nucleic acids. The nucleic acid extractions were centrifuged at 13,000 rpm for 20 min at 4°C. The supernatants were carefully removed and discarded, avoiding the white precipitate. The pellets were washed by adding 500 µL cold 70% ethanol, centrifuging at 13,000 rpm for 15 min at 4°C and removing the ethanol. The tubes were re-spun to remove any excess ethanol. The pellets were air dried for a few min and re-suspended into 100 µL RNase free water. The RNA was cleaned up using a RNeasy column (Qiagen, following manufacturer's instructions) which included a DNase treatment step on-column method (Qiagen) The RNA was stored at -80°C, storing a

separate 5 μ L sample for analysis of the RNA quality and yield using the Agilent Bioanalyser NanoChip and Nanodrop spectrophotometer.

2.10 Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) is a technique used to study protein-DNA interactions. It involves crosslinking all protein-DNA complexes, precipitating the protein of interest with an appropriate antibody and analysing the DNA fragments purified, typically through qPCR or high-throughput DNA sequencing. To perform the crosslinking, cultures for S417 and S418 were cultured in YEME-10 (with glycerol as a carbon source) to mid-late exponential phase ($OD_{450} = 1.5-2$). 37% formaldehyde solution was added to a final concentration of 1% before incubation for a further 20 min at 30 °C. To quench the remaining formaldehyde, glycine was added to a final concentration of 0.5 M before incubation for 5 min at room temperature. For each immunoprecipitation, 70 mL aliquots of cells were harvested by centrifugation at 6,000 x g for 2 min at 4 °C, washed twice with 25 mL ice-cold PBS and cell pellets were stored at -80 °C until required.

Keeping frozen in liquid nitrogen, pellets were cryogenically ground for 3 x 90 s before resuspension in 2.2 mL IP buffer (with 0.1 mg/mL RNase A + protease inhibitor). Cells were further disrupted, and DNA fragmented by sonication using a Diagenode Bioruptor. To do this, each pellet suspension was divided into 6 x 400 μ L aliquots and sonicated for 35 cycles of sonication (30 s on, 30 s off) at 4 °C. Samples were centrifuged at 16,000 x g for 30 min to remove cell debris and the supernatant transferred to a new tube. To visualise the degree of DNA fragmentation, a 20 μ L sample of the supernatant was treated with 2 μ L Riboshredder RNase Blend at room temperature for 30 min followed by 2 μ L proteinase K at 55 °C for 2 h and de-crosslinked at 65 °C for 6 h. This sample was run on a 1.2% agarose gel to check DNA fragment size; a smear ranging from ~150 bp to 600 bp was seen and judged suitable for sequencing. To perform the immunoprecipitation, the chromatin samples were first pre-cleared with the appropriate Protein A/G magnetic beads; this is a process that reduces non-specific binding by removing proteins that may bind immunoglobulins. Protein G magnetic beads (New England Biolabs) were used for RNAP and 3XFLAG immunoprecipitations. Pre-clearing was performed with 50 μ L magnetic beads incubated at 4 °C for 1.5 h rotating at 20

rpm. Following incubation, the magnetic beads were immobilised with a magnetic rack and the cleared supernatant was transferred to a new tube. 100 μ L of this pre-cleared chromatin was removed and stored at -20 °C as an "input" control. 1–5 μ g of the following antibodies was added to the remaining cleared chromatin and incubated at 4 °C for 2.5 h rotating at 20 rpm:

- 7 μ L anti-FLAG M2 monoclonal antibody (Sigma F18041MG)

30 μ L of appropriate magnetic beads were added and incubated for a further 1.5 h. The beads were immobilised with a magnetic rack and the spent supernatant was discarded. The beads were gently resuspended in 750 μ L ice-cold IP buffer, transferred to a new tube and incubated at 4°C for 10 min rotating at 20 rpm. The same wash procedure was performed with 1 mL ice-cold IP + salt buffer, IP wash buffer and finally pH 8 TE buffer. After washing, the magnetic beads were resuspended in 100 μ L 1 x IP elution buffer + 5 μ L RiboShredder RNase blend and incubated at room temperature for 30 min. At this point, the "input" control sample was thawed and incubated with 5 μ L RiboShredder RNase blend at RT for 30 min before de-crosslinking both the IP and input samples in a 65°C water bath overnight. The following day the beads were immobilised and washed with an additional 50 μ L pH 7.5 TE buffer. 50 μ L pH 7.5 TE buffer was added to the input sample. 150 μ L IP and input samples were finally purified through a Qiagen MinElute PCR purification kit and eluted with 22 μ L MQ water.

2.10.1 qPCR

qPCR was performed in a 96-well PCR plate using a StepOnePlus Real-Time PCR system (Applied Biosystems). A dilution series for genomic M145 DNA was prepared, ranging from 30-3,000,000 copies. The primers were prepared to a 1/20 dilution and the input control DNA was diluted to 1/1000, immunoprecipitated DNA in 1/10 dilution. A 12X GoTaq qPCR master mix was prepared for each primer set. 23 μ L of the master mix was added to each well followed by 2 μ L of DNA or water. Data was collected and analysed using Applied Biosystems® Real-Time PCR Software.

2.11 S1 nuclease protection assay

2.11.1 Preparing the gel

The glass plates and comb were washed with detergent and polished with 70% ethanol. The glass plates were mounted onto the stand. The gel was made using 16 mL UreaGel concentrate, 29 mL UreaGel diluent, 5 mL UreaGel buffer, 400 μ L 10% APS and 20 μ L TEMED to a total volume of 50 mL. The gel was poured into the mould making sure not to avoid bubbles. The comb was added, and the stand was rotated to position the glass plates vertical. The gel was left to polymerise for 1 h and thereafter the comb was removed. The gel-glass plates were placed into the gel tank and 1X TBE was poured gently into the wells to remove any bubbles, additionally a syringe was used to remove any urea from the wells. The pump in the gel tank was switched on and set to 55°C and the gel was pre-run for 45 mins. Finally 3 μ L samples were loaded.

2.11.2 S1 nuclease protection assay

To create the probe, the reverse primer (downstream from the 5' end) was first labelled with 32 P as follows: 2 μ L oligonucleotide (5 pmol/ μ L), 4 μ L 10 X T4 polynucleotide kinase buffer, 5 μ L ATP [γ - 32 P], 1 μ L T4 polynucleotide kinase and made up to 40 μ L with RNase free water. Then incubated for 30 min at 37°C and heated to 80°C for 10 min followed by cooling on ice. The oligonucleotide was purified using a Sephadex matrix. The Sephadex column was placed into a clean 1.5 mL microcentrifuge tube and centrifuged at 1000 x g for 1 min at RT. The eluted buffer was discarded and placed into a clean tube again. 300 μ L RNase free water was placed into the column and centrifuged at 1000 x g for 2 min at RT. The oligonucleotide was mixed carefully and applied to the prepared Sephadex column and centrifuged at 1000 x g for 4 min at RT, this purified column was stored at -20°C.

To perform S1 mapping, 30 μ g RNA was mixed with 5-10 ng of probe and lyophilised with a SpeediVac. The dried pellet was re-suspended in 20 μ L 1X S1 hybridisation buffer, incubated in a water bath at 70 °C for 10 min before allowing the temperature to fall to 42 °C overnight. 300 μ L S1 digestion mix was forcibly added to the mixture, thoroughly mixed and placed in ice water until all samples were ready. The digest mix was incubated in a water bath at 37 °C for

45 min until the reaction was stopped by addition of 75 μ L S1 stop buffer. 1 μ L glycogen and 400 μ L isopropanol was added, mixed and incubated at -20°C for 1 h. The mixture was centrifuged at 16,000 x g at 4°C for 15 min and all supernatant was removed. The pellet was washed with 70% ethanol and air-dried for 10 min. The pellet was then re-suspended in 6 μ L S1 formamide loading buffer, heated at 95 °C for 2 min and 3 μ L was run on an 8% UREA-PAGE sequencing gel at 600 V for 1 h 20 min with 1 X TBE running buffer. The gel was dried onto Whatman paper and exposed overnight in a phosphorimaging cassette before visualisation the next day with a Typhoon phosphorimager.

Results chapter I: Investigating the
role of the RbpA basic linker in
Streptomyces coelicolor A3 (2)

3.1.0 Overview

The RbpA-Sigma Interaction Domain (SID) has been the most extensively studied region of RbpA and has been shown to interact with the σ_2 domain of the principle sigma factors in both *S. coelicolor* and *M. tuberculosis*. However the functions of the N-terminal tail, RbpA core domain (RCD) and basic linker (BL) are less well defined. This chapter focuses on elucidating the function of the RbpA basic linker in *S. coelicolor*.

At the outset of this investigation, mutagenesis had revealed that the RbpA BL residue R80 played an important role in RbpA function *in vivo* and *in vitro*. An X-ray crystal structure and subsequently cryo-EM structural studies positioned the RbpA BL to interact with the extended -10 promoter element (Boyaci et al., 2018; Hubin et al., 2015; Hubin et al., 2017a) (Fig 3.1). The positively charged R79 residue in *M. tuberculosis* (RbpA^{R80} in *S. coelicolor*) is positioned closely to interact with the negatively charged positions -13 and -14 on the DNA backbone (Hubin et al., 2017a). Further, truncations of RbpA lacking the BL-SID (residues 1-79) were unable to activate transcription *in vitro*, suggesting that key elements for activation are in the BL-SID (Hubin et al., 2015). Formaldehyde cross-linking analysis using full length RbpA and BL-SID truncations suggested a DNA binding role (Hubin et al., 2015). However, the roles of the three conserved lysines in the BL (K74, K75 and K77) were less clear. Preliminary work (A. Tabib-Salazar and M. Paget, personal communication) revealed that mutants lacking all three lysine residues fully restored normal growth to a $\Delta rbpA$ mutant, while such a mutant could not cross-link DNA nor activate transcription *in vitro*. Although apparently contradictory, these data suggest a role for the lysine BL residues, possibly overlapping with that of R80. Residues K74, K75 and K77 and R80 (*S. coelicolor*) are very well conserved across the actinobacteria and therefore suggestive of an important role for RbpA-BL in the overall function of RbpA, possibly DNA binding or enhancing transcription activation. Therefore, the first section of this chapter investigates the phenotypes of the RbpA BL mutant alleles in the $\Delta rbpA$ conditional mutant S401 using solid and liquid media. The second section of this chapter uses RNA-seq to identify the changes in the *S. coelicolor* transcriptome as a direct result of the RbpA BL mutants.

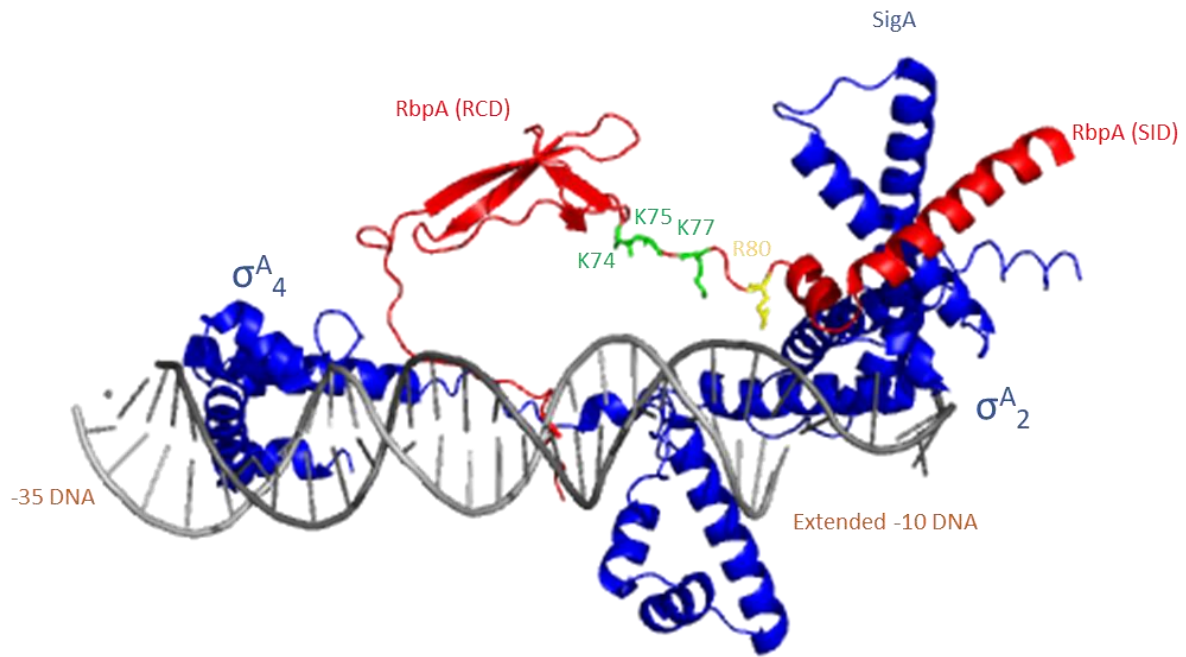


Figure 3.1. Cryo-EM structure of RNAP- σ^A holoenzyme with RbpA and complexed with upstream fork DNA in *M. tuberculosis*. The RNA polymerase subunits, β , β' , α_1 , α_2 , ω and Fdx have been removed from the structure to focus on the position of the RbpA BL. RbpA in red and σ^A in blue, nt-DNA in dark grey and t-DNA in light grey. The basic linker residues in RbpA, lysines K74, K75 and K77 in green and arginine R80 in yellow. The structure was downloaded from RCSB PDB, accession number 6C04 and visualised with PyMOL (version 2.0.7; Hubin, et al., 2017a).

3.1.1 Construction of the basic-linker mutants in *S. coelicolor*

A $\Delta rbpA$ mutant strain in *S. coelicolor* exhibits a small colony morphology, because of its paralogous *rbpB*. RbpB is not essential in *S. coelicolor* and a $\Delta rbpB$ mutant does not have a growth phenotype. However, a $\Delta rbpA \Delta rbpB$ mutant strain fails to grow, therefore all experiments were performed in *rbpA* null mutant backgrounds in which the paralogous *rbpB* was also knocked out as it might mask the effects of certain alleles (A. Tabib-Salazar and M. Paget, unpublished).

The RbpA-BL mutants were introduced into S401 strain (A. Tabib-Salazar and M. Paget, unpublished), in which the only copy of *rbpA* is induced by thiostrepton. In this background the phenotypes of mutants can therefore be fully established when cultured in the absence of

thiostrepton. As previously (Tabib-Salazar et al., 2013), alanine substitution was used to minimize unwanted structural effects.

The starting plasmid for constructing the mutations was pMT3000::*rbpA-NdeI* (created by Aline Tabib-Salazar), which includes 150 bp *rbpA* upstream DNA and an *NdeI* restriction site inserted at the start codon of the *rbpA* gene, to enable sub-cloning into expression vectors. The pMT3000::*rbpA-NdeI* was previously shown to fully complement an Δ *rbpA* mutant (Tabib-Salazar et al., 2013). This plasmid was used as a template to engineer the BL mutant PCR products, *rbpA*^{K74A}, *rbpA*^{K77A}, *rbpA*^{R80A}, *rbpA*^{3KA} (K74A, K75A and K77A) and *rbpA*^{3KRA} (K74, K75, K77 and R80 substituted with alanine) using inverse PCR and confirmed by sequencing.

PCR primers were designed to insert flanking *HindIII* sites 135 bp upstream of the *rbpA* start codon and immediately upstream of the stop codon, allowing the creation of an in-frame fusion with the 3xFLAG epitope tag in the integrative plasmid pRT802::3XFLAG (Methods, Table 2.1). pRT802, like pIJ6902 (integrates at the ϕ C31 site), integrates in the genome of *S. coelicolor*, in this case at the ϕ BT1 attachment site (Gregory et al., 2003). PCR products for the *rbpA*^{WT}, *rbpA*^{K74A}, *rbpA*^{K77A}, *rbpA*^{R80A}, *rbpA*^{3KA} and *rbpA*^{3KRA} containing the engineered *HindIII* sites were cloned initially into pBluescript II SK+, checked for SNPs by DNA sequencing, and then subcloned as *HindIII* fragments into pRT802:: 3XFLAG tag, where fusions and between the *rbpA* reading frame and the 3XFLAG tag were confirmed using *SpeI*/*BamHI* diagnostic digests. Following this, recombinant plasmids were transformed into *E. coli* ETR competent cells followed by conjugation into *S. coelicolor* S401 selecting for the incoming plasmid (Km^R) and supplementing with thiostrepton to ensure that the WT *rbpA* was expressed (Fig. 3.2). Thiostrepton was included in all sub-culturing media to reduce the incidence of suppressor development.

3.1.2 Phenotype of RbpA BL mutants in *S. coelicolor*

S. coelicolor S401 (pRT802::*rbpA*^{WT}) along with strains expressing the equivalent BL mutants (K74A, K77A, 3KA, 3KRA) or vector only control, were plated to single colonies on MS and SMMS agars to test for the effects of the mutagenesis on morphology and antibiotic production. Strains were cultured with/without thiostrepton for 4-5 days at 30°C (Fig. 3.3). As expected S401 (pRT802::*rbpA*^{WT}) grew equally well in the presence and absence of

thiostrepton because there was at least one copy of *rbpA* expressed in either condition. Whilst, S401 (pRT802) formed very small colonies in the absence of the inducer which is consistent with an S101 ($\Delta rbpA$) background. Consistent with previous observations, S401 (pRT802::*rbpA*^{3KRA}) fully complemented the S401 -Thio phenotype on MS and SMMS agar (Fig. 3.3-3.4), confirming that the three lysine residues are not required for normal growth *in vivo*, despite being required to activate transcription *in vitro*. Strains expressing RbpA^{R80A} and RbpA^{3KRA} produced a considerable amount of ACT but grew significantly better than S401 (pRT802) in the absence of thiostrepton. These data indicated that although RbpA appears to be essential, the BL itself is not essential for growth. Interestingly, although expression of RbpA^{R80A} leads to elevated ACT production, the additional substitution of all of the lysine residues (RbpA^{3KRA}) displays a more profound effect on ACT production, both on MS agar and SMMS; this is the first indication of an *in vivo* role for the lysine residues, perhaps in partnership with R80.

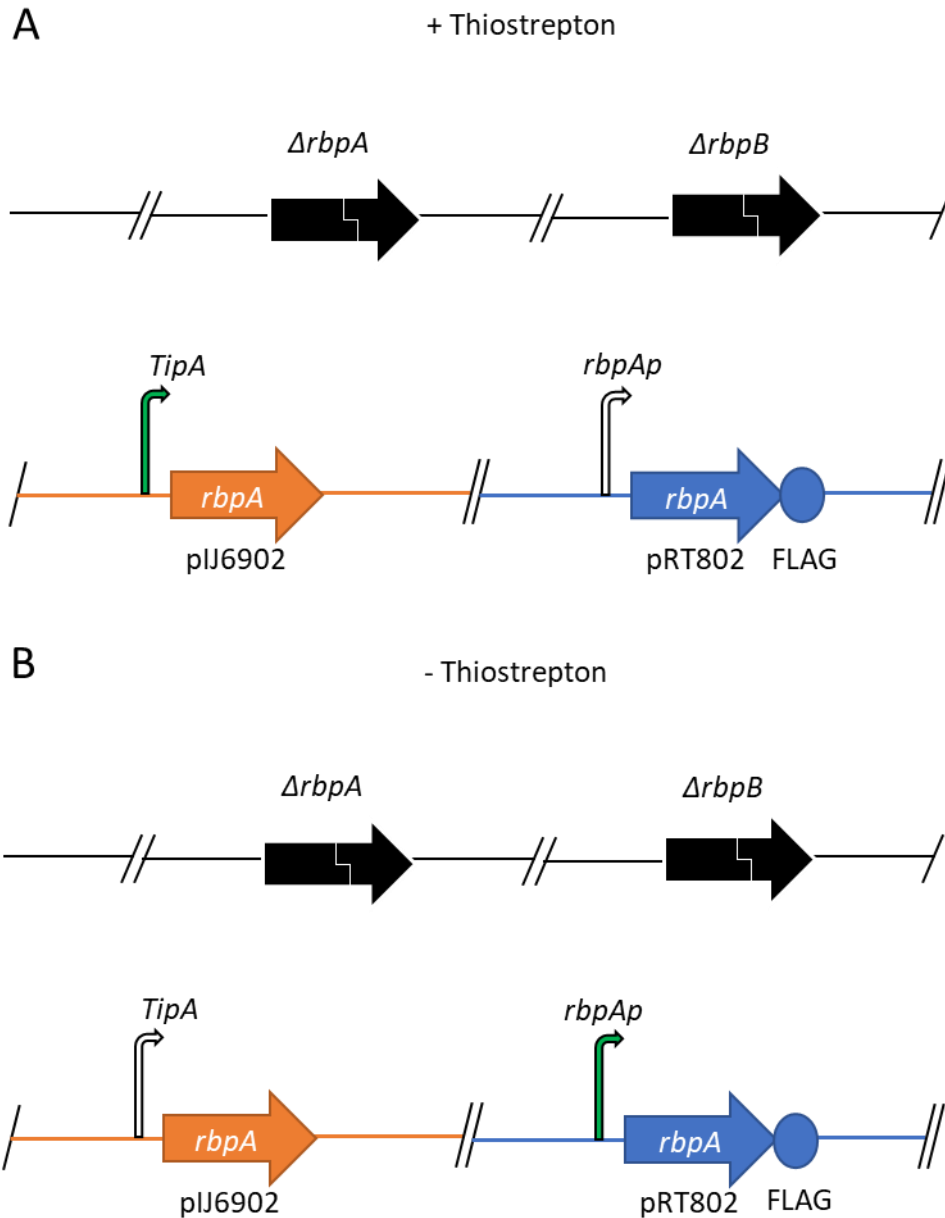


Figure 3.2. Engineering of the RbpA BL mutants in S401. S401 pIJ6902 (orange) integrated at ϕ HC31 site and the *tipA* promoter is dependent on the inducer thiostrepton. pRT802::FLAG (blue) integrates at the ϕ attP site and uses the native *rbpA* promoter. (A) Presence of thiostrepton activates *tipA* promoter (green) and thus expresses *rbpA*. (B) In the absence of thiostrepton *rbpA* in pIJ6902 is not expressed, allowing pRT802::*rbpA* and the *rbpA* BL mutant allele phenotypes to be investigated. The green filled in arrows represent the copy of *rbpA* being expressed.

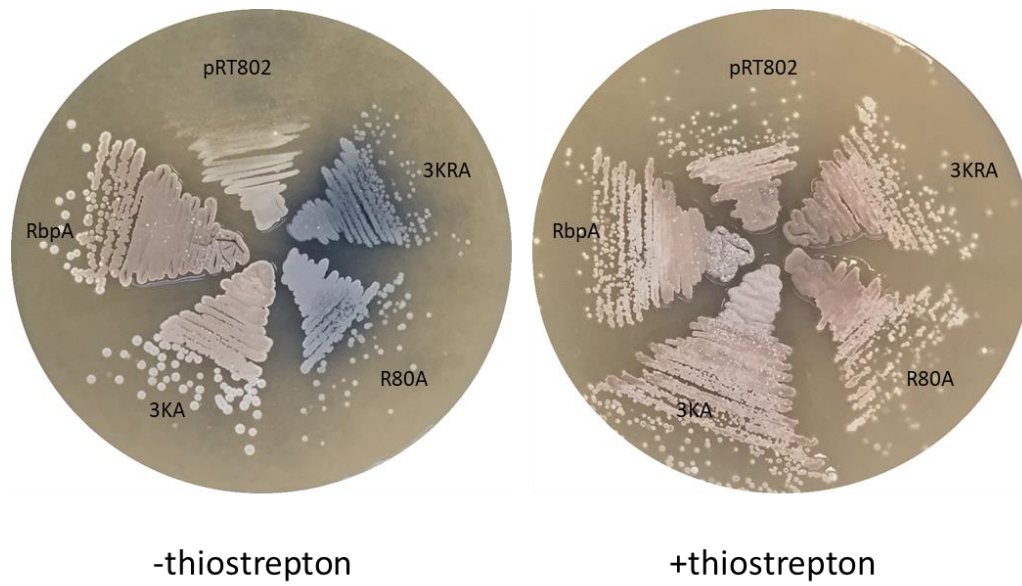


Figure 3.3. Phenotype of the RbpA BL mutants on MS agar. S401 (pRT802::*rbpA^{WT}*) (RbpA), S401 (pRT802) (pRT802), S401 (pRT802::*rbpA^{R80A}*) (R80A), S401 (pRT802::*rbpA^{3KA}*) (3KA) and S401 (pRT802::*rbpA^{3KRA}*) (3KRA) have been plated onto MS agar +/- 15 µg/mL thiostrepton and incubated at 30°C for 4-5 days.

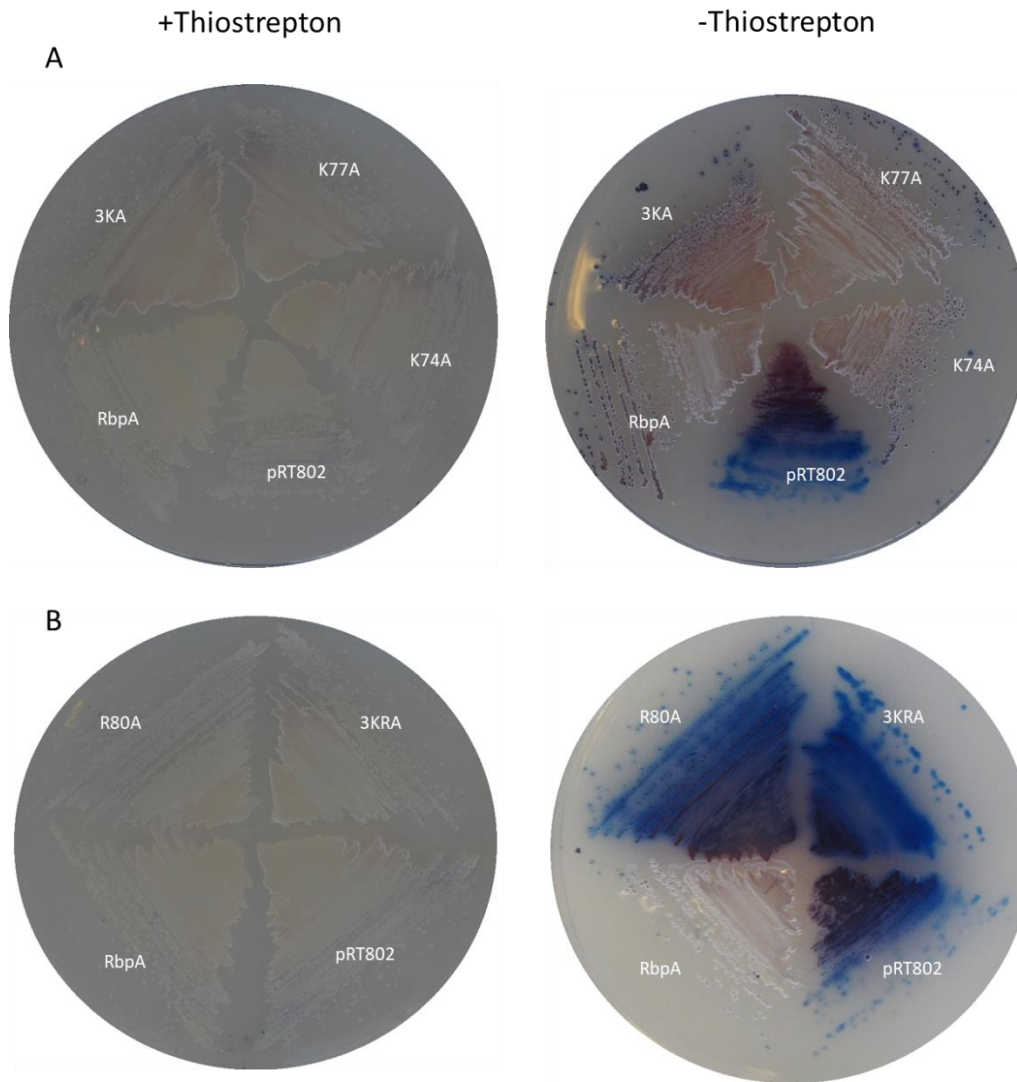


Figure 3.4. The phenotype of RbpA BL mutants on SMMS agar. (A) S401 (pRT802::*rbpA*^{3KA}) (3KA), S401 (pRT802::*rbpA*^{K77A}) (K77A), S401 (pRT802::*rbpA*^{K74A}) (K74A), S401 (pRT802) (pRT802), S401 (pRT802::*rbpA*^{WT}) (RbpA). (B) S401 (pRT802::*rbpA*^{WT}) (RbpA), S401 (pRT802) (pRT802), S401 (pRT802::*rbpA*^{R80A}) (R80A) and S401 (pRT802::*rbpA*^{3KRA}) (3KRA) have been plated onto SMMS agar – /+ (right and left respectively) 15 µg/mL thiostrepton and incubated at 30°C for 4-5 days.

3.1.3 Growth curves for the BL mutants in *S. coelicolor*

The BL mutants show well defined phenotypes on solid media therefore this was investigated in further detail by culturing in liquid media. The strains were cultured in an enriched YEME liquid media to test the difference in the rate of growth between the RbpA BL mutants. No thiostrepton was added to the cultures to prevent expression of the chromosomal conditional *rbpA* allele, kanamycin was included in the media to support stable expression of the BL mutant alleles in pRT802. Samples for optical density (OD₄₅₀) readings were taken over a period of 44 h (Fig. 3.5).

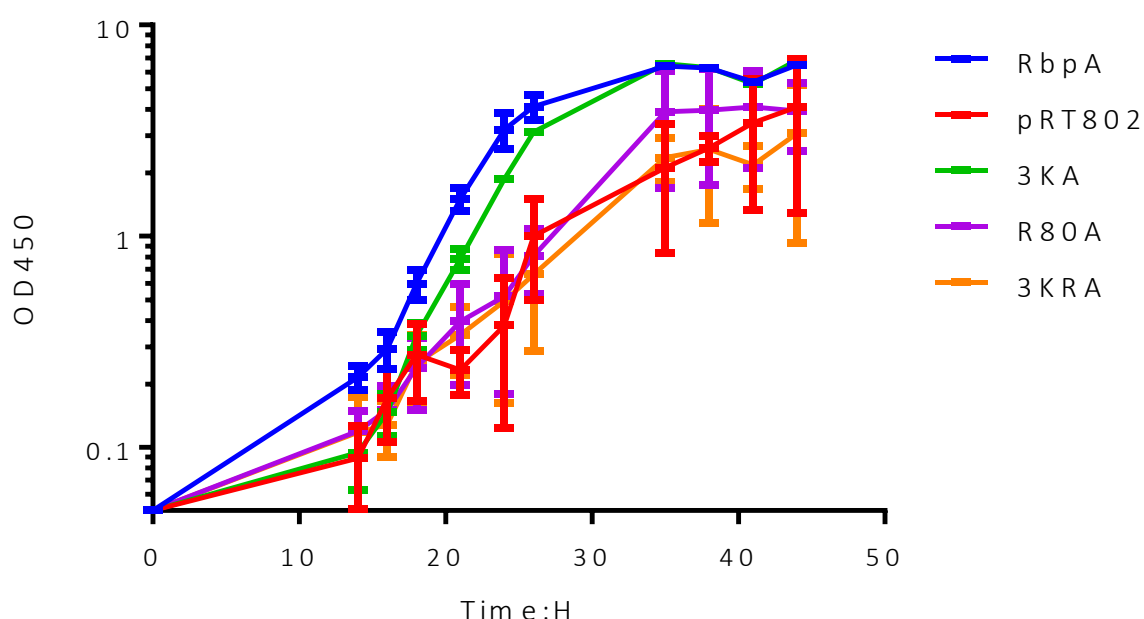


Figure 3.5. Growth curves of the RbpA BL mutants in *S. coelicolor*. S401 (pRT802::*rbpA*^{WT}; blue), S401 (pRT802; red), S401 (pRT802::*rbpA*^{R80A}; purple), S401 (pRT802::*rbpA*^{3KA}; green) and S401 (pRT802::*rbpA*^{3KRA}; orange) were cultured in YEME supplemented with glycerol at 30°C, 300 rpm without addition of thiostrepton. The results represent three biological replicates for S401 (pRT802::*rbpA*^{WT}) and S401 (pRT802::*rbpA*^{3KA}). The results of two biological replicates shown for S401 (pRT802::*rbpA*^{R80A}) at 21, 24, 35, 38 and 41 h, S401 (pRT802::*rbpA*^{3KRA}) at 24, 26, 38 and 44 h and S401 (pRT802) at 24, 35 and 44 h. Error bars denote mean with standard deviation.

The growth curve using OD showed that the S401 (pRT802::*rbpA*^{3KA}) and S401 (pRT802::*rbpA*^{WT}) grew at a similar rate whereas S401 (pRT802::*rbpA*^{R80A}) and S401 (pRT802::*rbpA*^{3KRA}) showed an extended lag phase compared to the *RbpA*^{WT} and delayed entry into exponential phase. Despite this, both were still viable in liquid media, which supports the phenotype seen on solid media. It is difficult to draw many conclusions from this growth curve as the negative control, S401 (pRT802) grew at a similar rate to S401 (pRT802::*rbpA*^{3KRA}). During the growth curves, it was observed that the BL mutant flasks were changing the colour of the media with the production of secondary metabolites. Therefore, the levels of the blue pigment, actinorhodin was quantified. The *RbpA* BL mutants were pre-germinated for 3 h and cultured in RG-2 media nitrogen limited (Bystrykh et al., 1996). The detection assay for actinorhodin was performed to analyse the intracellular γ -actinorhodin. The cells were re-suspended in 1N KOH and incubated at room temperature for 1 h and centrifuged, the supernatant was collected and the A_{640} was measured (Fig. 3.6). The quantification of ACT showed that the S401 (pRT802) and S401 (pRT802::*rbpA*^{3KRA}) strains were the highest ACT producers.

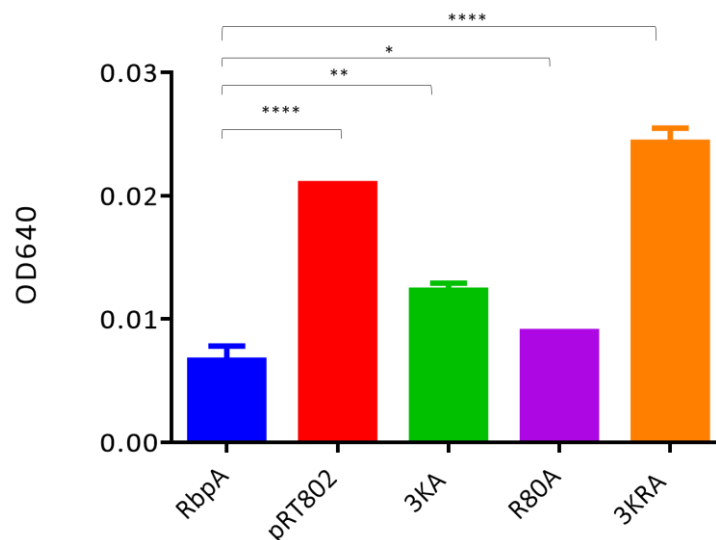


Figure 3.6. Actinorhodin production in the RbpA BL mutants. Spores for the BL mutants, (pRT802::*rbpA*^{WT}; blue), S401 (pRT802; red), S401 (pRT802::*rbpA*^{3KA}; green), S401 (pRT802::*rbpA*^{R80A}; purple) and S401 (pRT802::*rbpA*^{3KRA}; orange) were normalised and pre-germinated for 3 h following addition to RG-2 media containing no nitrogen and samples were collected at 72 h. This result is from one biological replicate (3 technical replicates) and the error bars denote mean with standard deviation. Unpaired T-tests: RbpA vs pRT802 p-value= <0.0001, RbpA vs 3KA p-value= 0.0016, RbpA vs R80A p-value= 0.249, RbpA vs 3KRA p-value=<0.0001.

3.1.4 Expression levels of the BL mutants in *S. coelicolor*

The growth curve and the solid media provide evidence of a phenotype for S401 (pRT802::*rbpA*^{R80A}) and S401 (pRT802::*rbpA*^{3KRA}) and a functional role of the RbpA BL in growth. However, it is possible that the more severe phenotype seen in these BL mutants is not decreased DNA interaction but rather decreased protein stability. To address this, this section investigates the expression levels of *rbpA* in each of the strains.

The spore stocks for S401 (pRT802::*rbpA*^{WT}), S401 (pRT802), S401 (pRT802::*rbpA*^{R80A}), S401 (pRT802::*rbpA*^{3KA}) and S401 (pRT802::*rbpA*^{3KRA}) were all 3XFLAG (Table 2.2) were normalised and cultured in enriched YEME media. At OD₄₅₀ 0.8, 10 mL samples were harvested (Methods; 2.4) and the protein concentrations of the cell lysate were quantified using the Qubit protein assay kit and each loading sample was subsequently normalised to 15 µg before running on a

10% SDS gel. The membrane was blotted for RbpA using the primary anti-FLAG antibody. The secondary antibody for RbpA was anti-mouse anti-HRP and the membrane was subsequently imaged using the LI-COR. The expression levels of the RbpA BL mutants were quantified relative to *rbpA^{WT}* using median signal values from the LI-COR. All three of the BL mutant alleles, *rbpA^{3KA}*, *rbpA^{R80A}* and *rbpA^{3KRA}* showed lower protein levels compared to the *rbpA^{WT}* (Fig. 3.7). Despite the lower expression levels of all three BL mutants, *rbpA^{3KA}* has the same colony morphology and growth rate as *rbpA^{WT}*. This indicates that the over-production of actinorhodin in *rbpA^{R80A}* and *rbpA^{3KRA}* may be due to the loss of the BL-DNA interactions, as the expression level of *rbpA^{3KA}* is consistent with the other BL mutants however does not produce actinorhodin. Therefore, the strains were taken forward for further experiments.

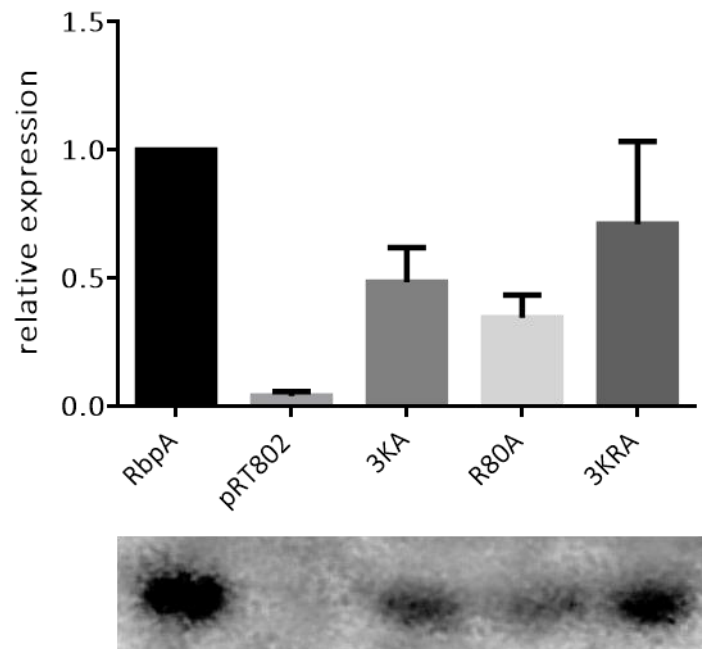


Figure 3.7. Expression levels of the RbpA BL mutant alleles in *S. coelicolor*. S401 (pRT802::*rbpA^{WT}*), S401 (pRT802), S401 (pRT802::*rbpA^{R80A}*), S401 (pRT802::*rbpA^{3KA}*) and S401 (pRT802::*rbpA^{3KRA}*) were cultured in YEME supplemented with glycerol to OD₄₅₀ 0.8, 10mL samples were harvested and prepared for western blot analysis. This data is compiled from 3 biological replicates and the expression of the BL mutants were analysed relative to *rbpA^{WT}*. Error bars denote mean with standard deviation.

3.1.5 Susceptibility of the RbpA BL mutants to rifampicin

Rifampicin has been used as a first-line drug for the treatment of active tuberculosis disease caused by *M. tuberculosis* since 1968 (Campbell et al., 2001). Rifampicin targets RNA polymerase, specifically the β subunit encoded by *rpoB* and may prevent the entry of elongating RNA into the active site of RNA polymerase (Campbell et al., 2001). The $\Delta rbpA$ strain exhibits a 15-fold lower MIC ($\mu\text{g mL}^{-1}$) to rifampicin compared to *rbpA*^{WT}, this increase in susceptibility to rifampicin is poorly understood (Newell et al., 2006). Therefore, this section investigates the effect of the BL mutant alleles on rifampicin resistance.

Spore dilutions containing 1 million spores per 10 μL for each of the RbpA BL strains was plated onto minimal media with increasing concentrations of rifampicin, 0-5 $\mu\text{g/mL}$ and incubated at 30°C for 3 days (Fig. 3.8). The *rbpA*^{WT} showed a gradual decrease in rifampicin resistance, whilst the $\Delta rbpA$ (pRT802) shows an increased sensitivity to rifampicin even at the lower concentrations (1 $\mu\text{g/mL}$) of rifampicin, which is consistent with previous findings. The S401 (pRT802::*rbpA*^{3KA}) strain demonstrates a similar profile to pRT802::*rbpA*^{WT} up to 2 $\mu\text{g/mL}$. The S401 (pRT802::*rbpA*^{R80A}) and S401 (pRT802::*rbpA*^{3KRA}) show sensitivity to rifampicin at very low concentrations, 0.5 $\mu\text{g/mL}$ and share a similar sensitivity profile to $\Delta rbpA$ (pRT802). This would suggest that the mutant allele's *rbpA*^{R80A} and *rbpA*^{3KRA} are implicated in rifampicin resistance of *S. coelicolor*. Alternatively, the $\Delta rbpA$ strain is more susceptible to rifampicin because of less active RNA polymerase. RbpA could be important for recovery of the cell from the stress caused by rifampicin. However, it was difficult to conclude that RbpA may require a fully functional BL to resist rifampicin without protein expression levels post rifampicin treatment. Therefore, western blot analysis was implemented to investigate the effect of rifampicin treatment on the expression levels of the BL mutants. However, rifampicin is capable of activating *sigR*, which goes onto activate the SigR regulon. RbpA is a member of the SigR regulon and thus is induced by rifampicin, therefore the western blots to investigate the RbpA^{WT} and RbpA BL expression levels upon rifampicin treatment were not presented here.

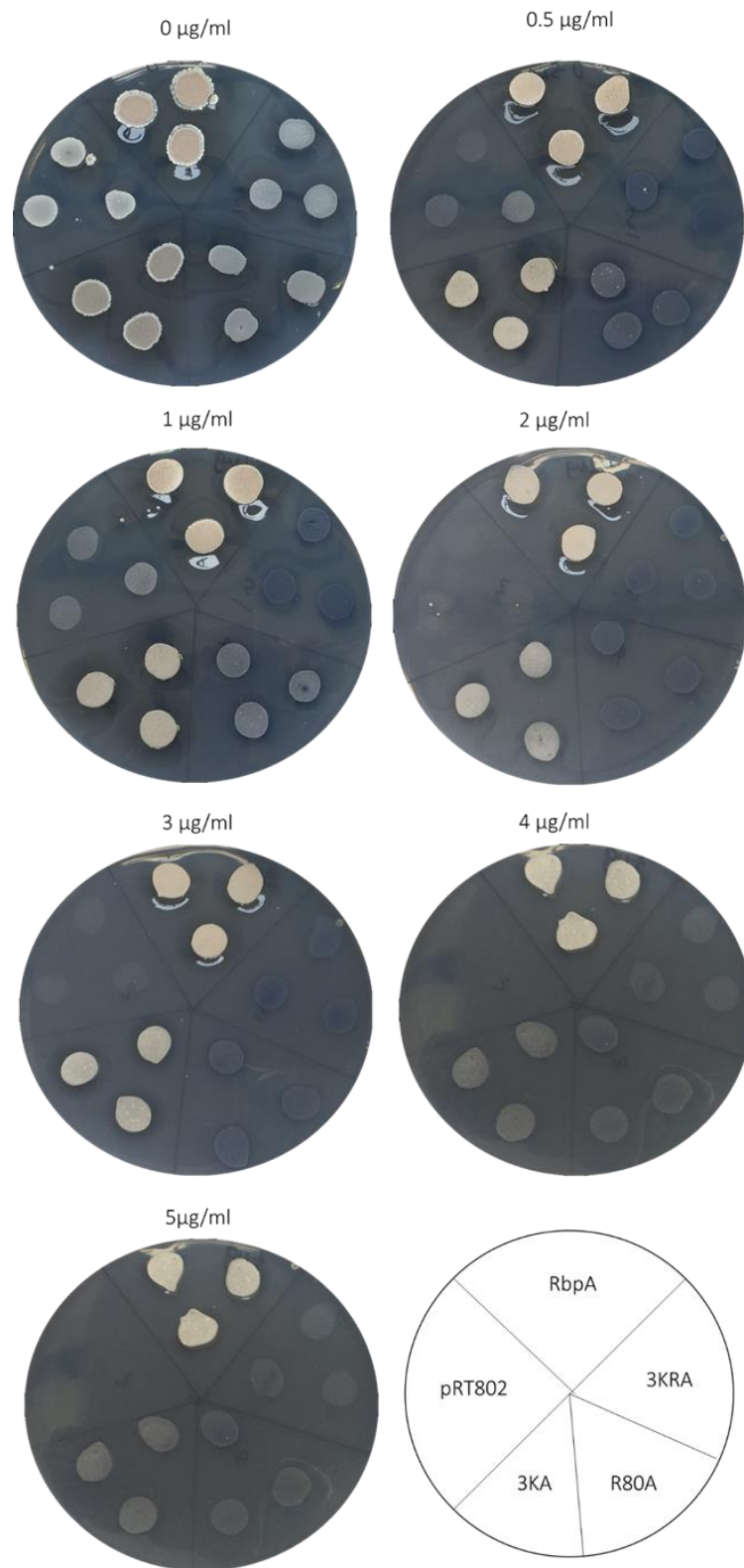


Figure 3.8. Sensitivity of the BL mutants in *S. coelicolor* to rifampicin. The spore stocks for S401 (pRT802::*rbpA*^{WT}), S401 (pRT802), S401 (pRT802::*rbpA*^{3KA}), S401 (pRT802::*rbpA*^{R80A}), S401 (pRT802::*rbpA*^{3KRA}) were normalised to 1 million spores per 10 µl and spotted onto minimal media with 0-5 µg/mL rifampicin. The plates were incubated at 30°C for 3 days.

3.1.6.1 RNA extractions in *S. coelicolor*

The *rbpA* BL mutant, S401 (pRT802::*rbpA*^{3KRA}) demonstrated increased production of ACT and reduced growth both on solid agar and liquid culture. Therefore, to investigate the significance of the S401 (pRT802::*rbpA*^{3KRA}) *in vivo*, the technique for RNA extraction was optimised so that RNA-seq analysis of the S401 (pRT802::*rbpA*^{WT}) could be compared to S401 (pRT802::*rbpA*^{3KRA}) and S401 (pRT802).

Traditionally *Streptomyces* RNA is prepared by flash freezing the cell pellet and grinding using a pestle and mortar (Jeong et al., 2016). This method was tested for the purpose of extracting RNA from S401 (pRT802::*rbpA*^{WT}). However, the method involved large time lags between times taken from centrifugation of cultures to addition of TRI reagent. There was also long periods of extensive handling and processing of the samples which increased the likelihood of RNase contamination. Therefore, an RNA extraction method from *M. tuberculosis* was modified for use in *S. coelicolor*. The spores for S401 (pRT802::*rbpA*^{WT}), S401 (pRT802) and S401 (pRT802::*rbpA*^{3KRA}) were normalised and pre-germinated for 3 h. The spores were cultured in 60 mL YEME supplemented with glycerol to OD₄₅₀ 0.8. At this point, 12 mL samples were added to stop solution at a 5:1 culture: stop solution followed by RNA protect and immediately centrifuged as a precautionary measure to prevent transcriptional changes during the handling of samples. The sample pellets were suspended in TRI reagent and ribolysed, the same protocol method of nucleic acid extraction, precipitation and RNA clean-up was followed as described in Methods; 2.6. The bioanalysis and nanodrop results all provided an indication that the RNA quality was good shown by the presence of two peaks for the 16S and 23S ribosomal subunits, 260/280 ratio of ~2 and 260/230 ratio of ~2 (Fig. 3.9).

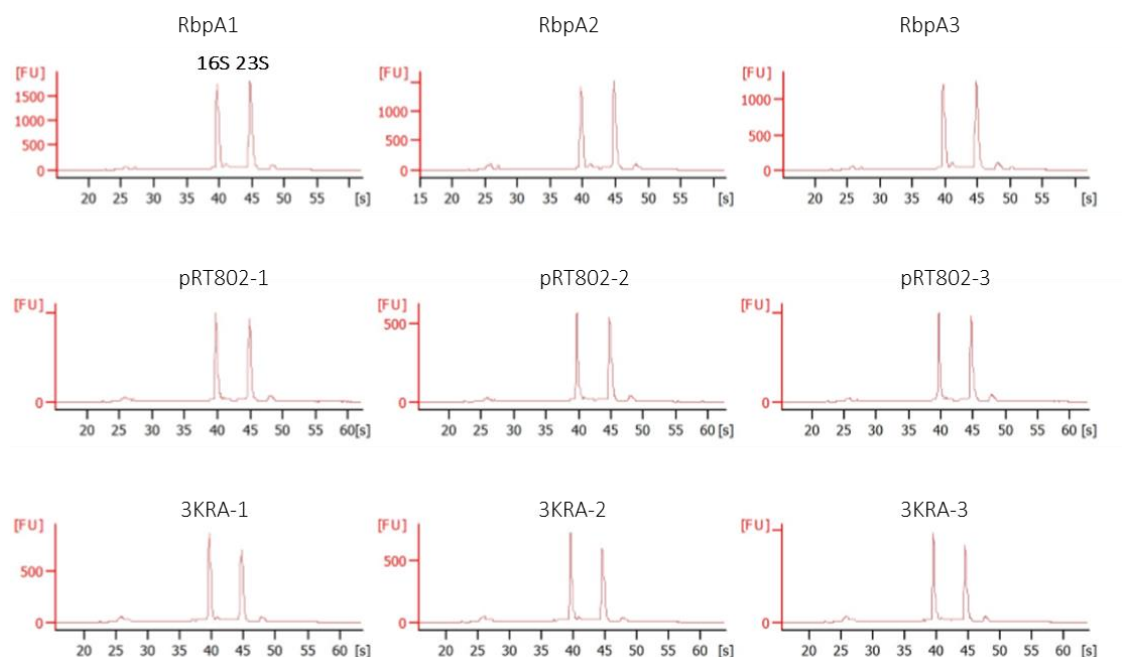


Figure 3.9. Bioanalyzer results of the RbpA BL RNA samples from *S. coelicolor*. The RNA samples were analysed using the Agilent 2100 Bioanalyzer NanoChip® systems. A-C S401 (pRT802::*rbpA*^{WT}), 3 biological replicates. D-F S401 (pRT802), 3 biological replicates. G-I S401 (pRT802::*rbpA*^{3KRA}) 3 biological replicates.

3.1.6.2 Sample preparation and the library preparation at Oxford genomics centre

The quality of the samples was checked at the Oxford genomic centre, Wellcome Trust for Human genetics (UK) by running the samples on a TapeStation and qubit for RNA quantification. The gel image (Appendix. 1) shows that all the samples had bright band for both ribosomal subunits, 16S and 23S, which is a good indication of the quality of the RNA samples. The library preparation of the 9 *S. coelicolor* samples (three biological replicates for S401 (pRT802::*rbpA*^{WT}), S401 (pRT802::*rbpA*^{3KRA}) and S401 (pRT802)) was commenced after the quality control checks were completed using the Truseq Stranded mRNA library prep kit from Illumina and the sequencer, Illumina HiSeq 4000. The library was prepared using 0.1- 4 µg total RNA, which was purified and fragmented to 150 bp to increase the mapping ability of the reads to the reference genome (Fig. 3.10). The first strand synthesis began with a SuperScript II reverse transcriptase which uses the RNA fragments randomly primed using random primers

to produce the first strand of complementary DNA. The second strand synthesis incorporates dUTP in place of the dTTP to produce a double stranded blunt ended cDNA molecule. The highly expressed ribosomal RNA, 16S, 23S and 5S, were removed from the total RNA samples, by using a ribodepletion kit (Ribozero Universal bacterial kit). This step ensures that the highly expressed ribosomal RNAs do not take up a high proportion of the sequencing reads and therefore the subtle changes in other genes can be analysed. This technique uses probe-hybridization to the ribosomal subunits which are removed using magnetic beads, the resulting supernatants are rRNA depleted. Indexing adapters were ligated to the double stranded cDNA to allow hybridisation of the cDNA to the flow cell. The cDNA was end repaired and the uridine was depleted. The double stranded cDNA which ligates the indexing adapters on both ends was enriched for use in the PCR, this was achieved by using a wide array of primers which anneal to the adapters and amplified both strands. The number of PCR cycles was kept to a minimum to prevent introducing errors. The subsequent library produced was validated, normalised and multiplexed. Multiplexing of the samples involved attaching unique bar codes to each cDNA molecule and pooling the samples into a single reaction, which produced millions of reads per sequence making this method quick and less expensive. Paired-end sequencing occurs over 100-300 bp, making the 5' and 3' sequenced ends of a poor quality generally but paired end sequencing increased sequence coverage by reading the sequence in both directions.

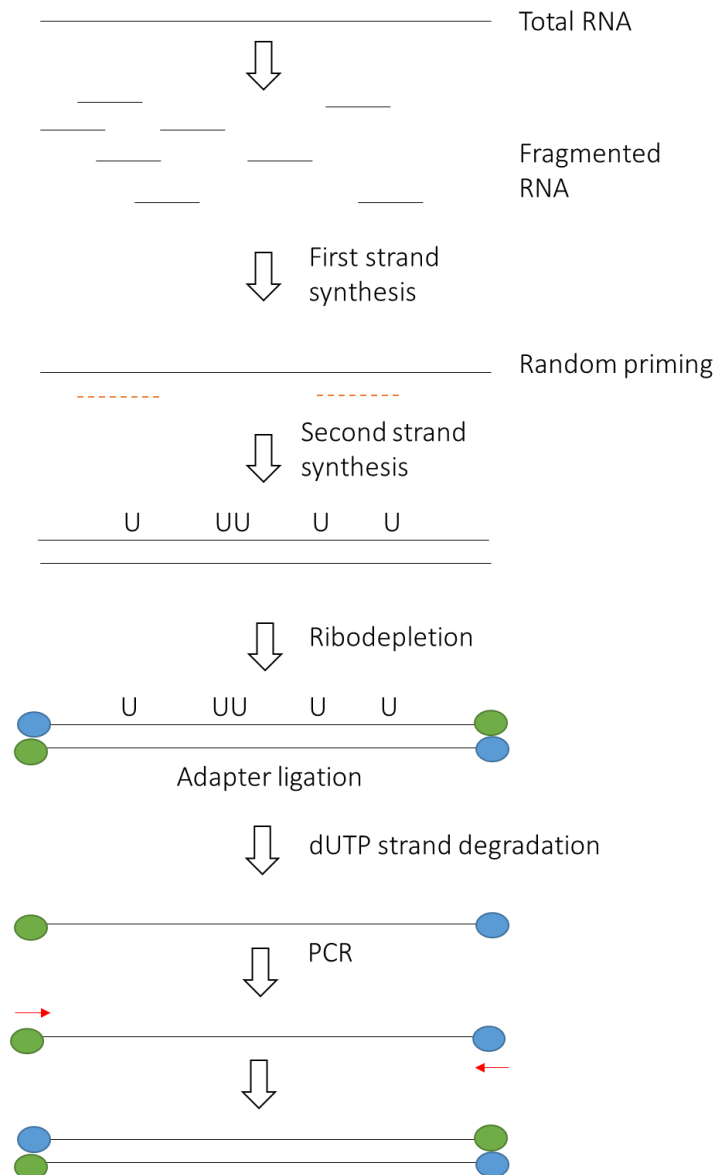


Figure 3.10. Overview of the library preparation carried out by the Oxford genomics centre. The RNA-seq library preparation was carried out by Oxford genomics centre, Wellcome trust for human genetics, UK. The Truseq Stranded mRNA library prep kit from Illumina was used to prepare the library. Total RNA was fragmented, and the first strand of cDNA was synthesised following random oligo priming. The second strand synthesis of cDNA was incorporated with dUTP. The ribosomal RNA was selectively removed using the Ribozero universal bacterial kit (NEB). Indexing adapters were added to the ends of the cDNA, and the cDNA was end repaired and uridine depleted. The samples were multiplexed, and PCR cycles were started.

3.1.7.1 RNA-sequencing bioinformatic analysis

RNA-seq requires the alignment of cDNA fragments to a reference genome, which minimises bias created during assays such as microarrays (Croucher & Thomson, 2010). RNA-seq is commonly used now for the study of transcriptomes because of its high sensitivity and accuracy. There are several pipelines that can be used to process raw sequencing files, this chapter focuses on the implementation of Cufflinks to obtain a differential gene expression list for the investigation of the importance of the RbpA BL (Fig. 3.12; Trapnell et al., 2011).

3.1.7.2 FastQC and trimming of the reads

The High-Performance computer (HPC) cluster at the University of Sussex was used to carry out the bioinformatic analysis of the FASTQ files. Each FASTQ file was uploaded onto WINSNP (version 5.10.4) and subsequently onto the module. To improve the alignment across the genome it is important to remove the adapters, this is DNA sequences which are not complementary to the reference genome (Table 2.6). Next the quality of the reads from the HiSeq 4000 sequencer was visualised using FastQC (version 0.11.5). FastQC processed the samples in the FASTQ format and provided a preview of GC%, quality of the base reads and potential adapter contamination. The FastQC analysis of the samples showed one peak for GC content at 66% which correlates with *S. coelicolor*. The overrepresented sequences indicated no hits for the adapters. However, the adapter sequences were checked independently in each sample using Text manipulation (Grep; Galaxy Tool version 1.0.0), which showed some matches. Therefore, Scythe (version 0.993b) was used to remove the 3' adapter sequences which can give poor quality bases, Table. 3.1, shows very little contamination of the reads with the adapter sequences. Reads usually consist of a deteriorating quality at the 3'-end therefore the reads were trimmed using Sickle (version 1.29) which trimmed the reads at the 3' end to remove the poor base quality (Table 3.2). A genome alignment of the samples was carried out before and after running Scythe and Sickle, this showed that by quality controlling the reads, mapping ability increased by ~3% (Fig. 3.11).

Sample	Contaminated reads		Contamination rate (%)	
	PE1	PE2	PE1	PE2
RbpA (1)	413601	537690	0.016	0.020
RbpA (2)	358563	487059	0.014	0.018
RbpA (3)	413077	510781	0.015	0.018
pRT802 (1)	1055316	1236762	0.032	0.037
pRT802 (2)	793136	933858	0.026	0.031
pRT802 (3)	1711324	1869721	0.047	0.051
3KRA (1)	1785212	1927009	0.044	0.047
3KRA (2)	2475563	2713199	0.055	0.060
3KRA (3)	732426	2710286	0.034	0.071

Table 3.1. The number of reads contaminated with adapters. Scythe (version 0.993b) detected adapters at the 3' end of the reads for the RNA extracted from S401 (pRT802::*rbpA^{WT}*), S401 (pRT802), S401 (pRT802::*rbpA^{3KRA}*). PE1= Paired end 1 and PE2= Paired end 2

Sample	Paired reads kept	Paired reads discarded
RbpA (1)	25932388	430
RbpA (2)	25505589	454
RbpA (3)	27341400	631
pRT802 (1)	32630986	774
pRT802 (2)	29628027	634
pRT802 (3)	35755658	990
3KRA (1)	39911505	1152
3KRA (2)	44604080	1279
3KRA (3)	37469114	944

Table 3.2. The number of reads kept and discarded after trimming of the *S. coelicolor* RNA samples. The reads from samples S401 (pRT802::*rbpA*^{WT}), S401 (pRT802), S401 (pRT802::*rbpA*^{3KRA}) were trimmed using Sickle (version 1.29).

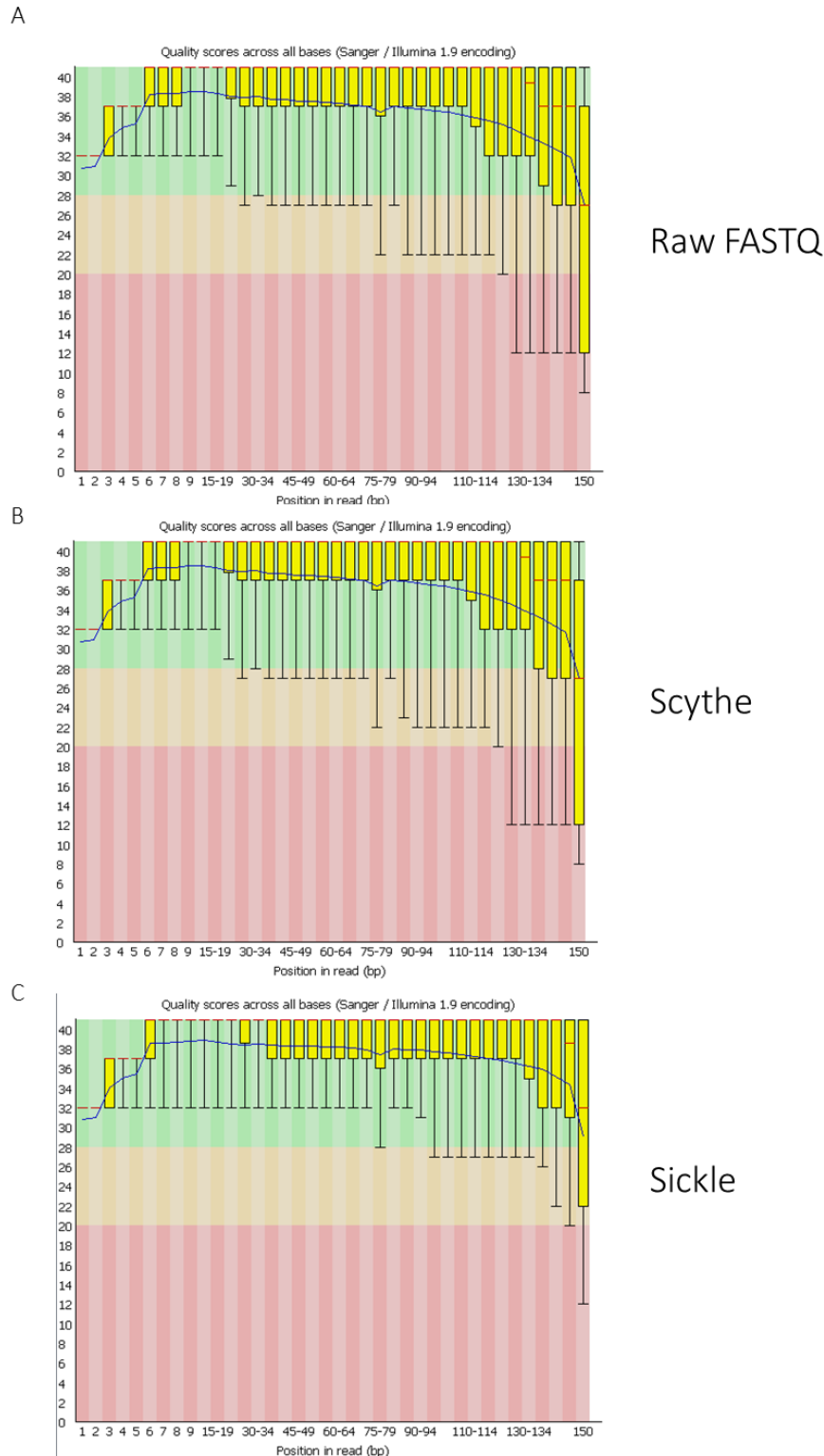


Figure 3.11. FastQC results of sample S401 (pRT802::*rbpA^{WT}*) (replicate-1, paired read 2) before and after trimming. FastQC (Galaxy tool, version 0.11.5) was used to check the quality of the reads. (A) Raw FASTQ file for S401 (pRT802::*rbpA^{WT}*) (1) (B) FASTQ file for S401 (pRT802::*rbpA^{WT}*) (1) processed using Scythe (version 0.993b) to remove any present adapters. (C) FASTQ file for S401 (pRT802::*rbpA^{WT}*) (1) trimmed using Sickle (version 1.29).

3.1.7.3 Mapping to the *Streptomyces coelicolor* A3 (2) genome

The reference genome, *Streptomyces coelicolor* A3 (2) NCBI: NC_003888.3 was downloaded from NCBI in GTF and FASTA format, the GTF file provides gene annotation whilst the FASTA file is the base sequence of the whole genome. The genome FASTA file was loaded in the HPC cluster using HISAT2 (version 2.1.0). HISAT2 is an alignment program which was used to map the sequencing reads to the *S. coelicolor* genome, this program has been developed from the HISAT and Bowtie2 genome builders. The trimmed paired end sample reads were aligned to the reference genome and the output file was in SAM format. The results of the alignment are summarised in Table 3.3, the mapped reads average at 95% to the *S. coelicolor* genome.

Sample	Number of reads	Mapped reads (concordant)	Mapped reads multiple times	Aligned pairs	Overall alignment rate
RbpA (1)	25932388	14891882 (57.43%)	71181 (0.27%)	3528201 (67.12%)	96.95%
RbpA (2)	25505589	14113062 (55.33%)	73207 (0.29%)	3903500 (67.86%)	96.68%
RbpA (3)	27341400	20056210 (73.35%)	663721 (2.43%)	1366847 (12.41%)	82.78%
pRT802 (1)	32630986	29905940 (91.65%)	145684 (0.45%)	899893 (26.95%)	96.31%
pRT802 (2)	29628027	26900751 (90.79%)	133442 (0.45%)	950592 (28.96%)	96.11%
pRT802 (3)	35755658	32600771 (91.18%)	191690 (0.54%)	1015888 (26.59%)	96.13%
3KRA (1)	39911505	36156628 (90.59%)	139236 (0.35%)	1399633 (34.86%)	96.77%
3KRA (2)	44604080	41035477 (92.00%)	178506 (0.40%)	1232479 (29.16%)	96.70%
3KRA (3)	37469114	34472583 (92.00%)	171913 (0.46%)	1013522 (28.43%)	96.65%

Table 3.3. Overview of the mapping ability of the sample reads to the *S. coelicolor* genome. The reads for S401 (pRT802::*rbpA*^{WT}), S401 (pRT802), S401 (pRT802::*rbpA*^{3KRA}) have been mapped to the *S. coelicolor* genome using HISAT2 (version 2.1.0) at a 95% overall alignment rate.

3.1.7.4 Assembling the transcriptome using Cufflinks

The mapped reads to the *S. coelicolor* genome using HISAT2 produced SAM files which were converted into their binary BAM files using SAMtools (version 0.1.19) and the subsequent BAM file was sorted using the same index and aligned to the reference genome. Cufflinks (version 2.2.1) is a suite of tools which assemble, merge, quantify and measure the differences in two conditions (Trapnell et al., 2011). The Cufflinks (version 2.2.1) algorithm was implemented to assemble individual transcripts from each of the RNA-seq reads that are aligned with the reference genome. Next the transcripts from the biological replicates were merged to improve the assembly quality and to reduce the read depth. The FPKM (fragments per kilobase of transcript per million mapped reads) is calculated after the merger of the biological replicates against the reference genome. The cuffmerged files were loaded into Cuffquant to calculate the transcript abundances by using the merged.gtf and BAM files, the output file was named abundances.cxb. Next, Cuffdiff was used to compare two conditions, RbpA^{3KRA} vs RbpA^{WT} or RbpA^{WT} and pRT802. The comparison between RbpA^{3KRA} vs RbpA^{WT} produced a gene_exp.diff file which can be viewed in Microsoft excel. The gene list provided FPKM values from the Cuffmerge application of the three biological replicates, log2 fold change, uncorrected p-value and false discovery rate (FDR) adjusted p-value of the test statistic using Benjamini-Hochberg correction for multiple testing. The differential gene expression data can all be viewed in R studio (version 1.0.153) using the CummeRbund package (version 2.0.0), this provides the individual replicate FPKM values for each replicate which can be used to identify anomalous results which falsely inflated gene expression differences.

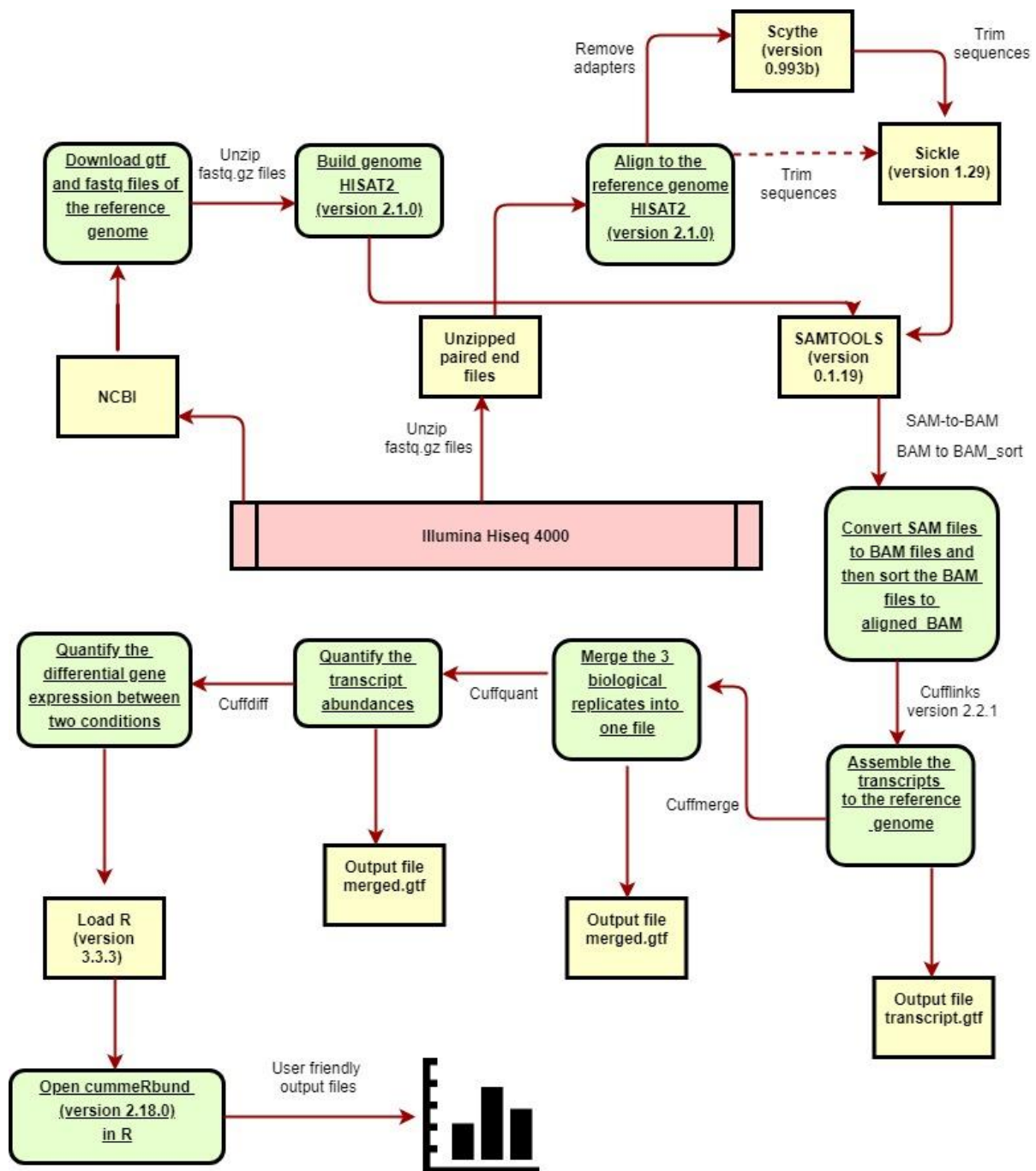


Figure 3.12. Overview of the RNA-sequencing pipeline. RNA-seq FASTQ files processed using the pipeline to obtain quantified differential transcript levels between RbpA^{WT} and RbpA^{3KRA} or RbpA^{WT} and pRT802 in *S. coelicolor*. Trackline arrow= alternative options, grey arrow= not a necessary step

3.1.8 Analysis of the differential gene expression between RbpA^{WT} and RbpA^{3KRA}

The differential gene expression between *rbpA*^{WT} and pRT802 (Δ *rbpA*) has been analysed in results chapter III 5.4.1, more extensively whilst this chapter focuses on the impact of *rbpA*^{3KRA} on gene expression identifying genes that were differentially expressed between S401 (pRT802::*rbpA*^{3KRA}) and S401 (pRT802::*rbpA*^{WT}). The up-regulated genes in S401 (pRT802::*rbpA*^{3KRA}) have also been discussed further in chapter III. The gene list was sorted according to significance (yes or no), resulting in 3788 significant genes (1432 up-regulated and 2356 down-regulated) out of 7069. CummeRbund was used to check the quality of the RNA-seq RbpA^{WT} and RbpA^{3KRA} samples. Boxplots showed that the log₁₀ (FPKM) was generally uniform between the three biological replicates (Fig. 3.13A). The dendrogram demonstrated that the three biological replicates for each sample, and RbpA^{3KRA} vs RbpA^{WT} cluster together based on the different genetic backgrounds (Fig. 3.13B). Lastly, the heatmap clustered 3277 genes (unable to load some of the genes) significantly different between *rbpA*^{3KRA} and *rbpA*^{WT} (Fig. 3.14) which showed a clear up-regulation and down-regulation of genes in both datasets.

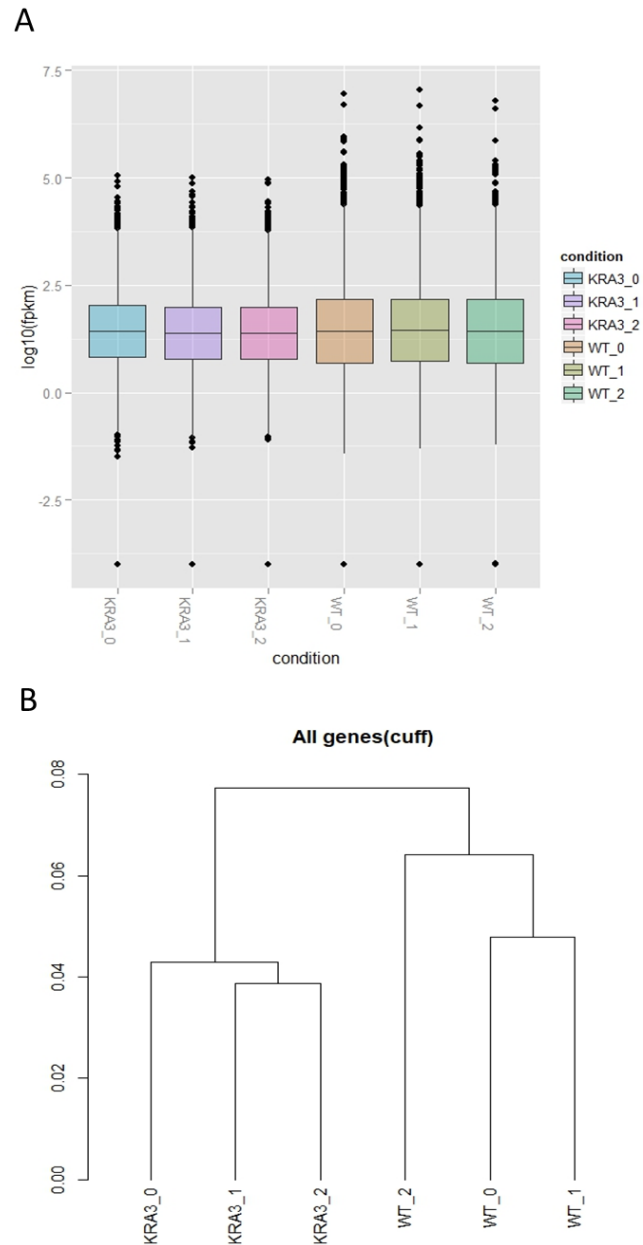


Figure 3.13. Quality of the RNA-seq comparing S401 (pRT802::*rbpA*^{WT}) and S401 (pRT802::*rbpA*^{3KRA}). (A) Boxplots show the uniformity of the biological replicates for S401 (pRT802::*rbpA*^{WT}) and S401 (pRT802::*rbpA*^{3KRA}). (B) Dendrogram shows the clustering of the biological replicates for S401 (pRT802::*rbpA*^{WT}) and S401 (pRT802::*rbpA*^{3KRA}).

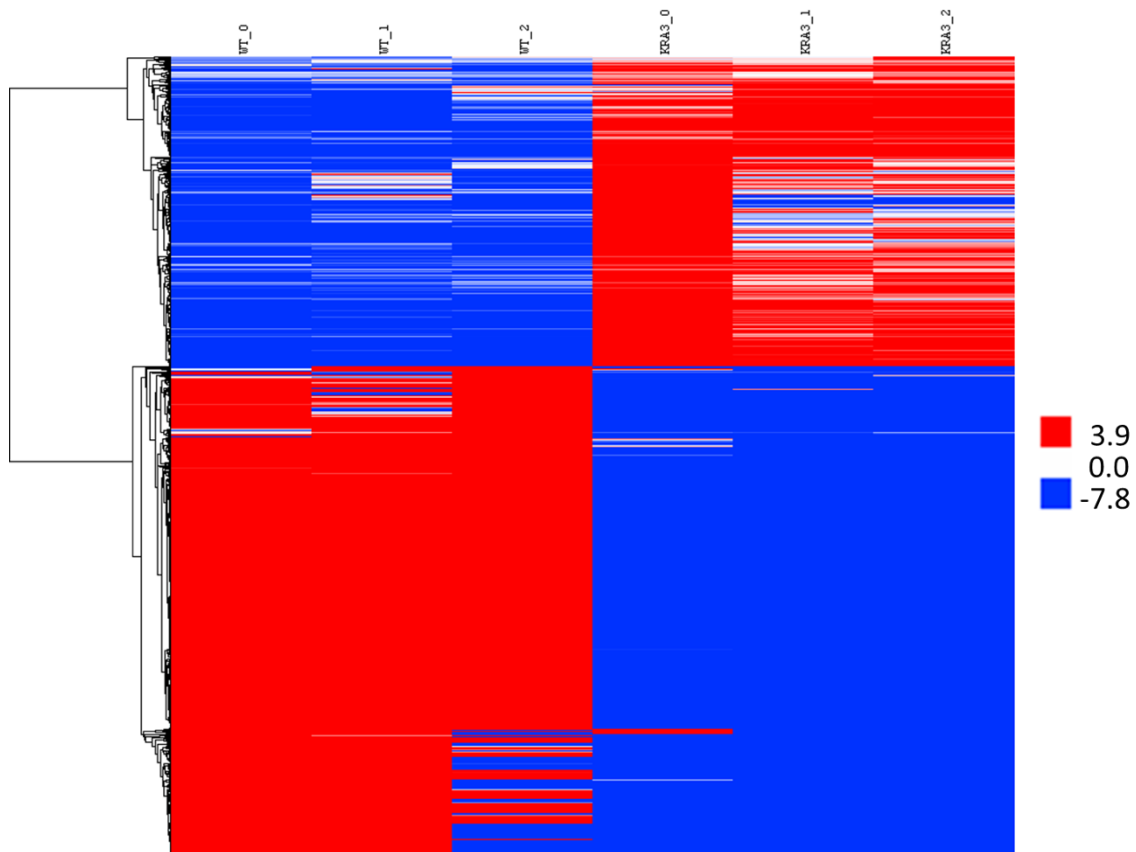


Figure 3.14. Heatmap generated using FPKM values for S401 (pRT802::*rbpA*^{WT}) and S401 (pRT802::*rbpA*^{3KRA}). The FPKM values for three biological replicates from the *rbpA*^{WT} and *rbpA*^{3KRA} samples were loaded into Gene Cluster 3.0. The genes were clustered, centred using the mean and average linked. The heatmap was viewed in Java TreeView (version 3.0). Red= up-regulated, blue =down-regulated. The colour bar represents the range of log2 (fold change).

3.1.8.1 Growth related genes are down-regulated in S401 pRT802::*rbpA*^{3KRA}

As S401 (pRT802::*rbpA*^{3KRA}) strain demonstrated a growth phenotype in both liquid and solid media, it was important to explore the transcriptome for growth-related genes to explain the reduced growth of S401 (pRT802::*rbpA*^{3KRA}). The RNA-seq data revealed that 46 ribosomal protein genes out of 52 were down-regulated by 1.48-135.2 fold in S401 (pRT802::*rbpA*^{3KRA}) compared to S401 (pRT802::*rbpA*^{WT}). Most interestingly, datasets comparing S401 (pRT802::*rbpA*^{3KRA}) and S401 (pRT802) revealed that 44 ribosomal protein genes are down-regulated in S401 (pRT802::*rbpA*^{3KRA}) and S401 (pRT802; Fig. 3.15).

Next, KEGG pathways were used to identify other groups of growth related genes. This revealed 20 amino acyl tRNA synthetase (out of a possible 29; (Fig. 3.16A) and 12 peptidoglycan biosynthesis genes (out of a possible 21; Fig. 3.16B) were down-regulated in S401 (pRT802::*rbpA*^{3KRA}) and S401 (pRT802). In addition to this, all RNAP subunits were also down-regulated in (pRT802::*rbpA*^{3KRA}) but not in S401 (pRT802; Fig. 3.17). It is unclear why the RNAP subunits are all down-regulated S401 (pRT802::*rbpA*^{3KRA}), it could suggest that global down-regulation of the growth related genes signals down-regulation of RNAP, however the lack of down-regulation in S401 (pRT802), disproves this theory. Nevertheless, the down-regulation of growth related genes supports the growth phenotype of the S401 (pRT802::*rbpA*^{3KRA}) shown in liquid media.

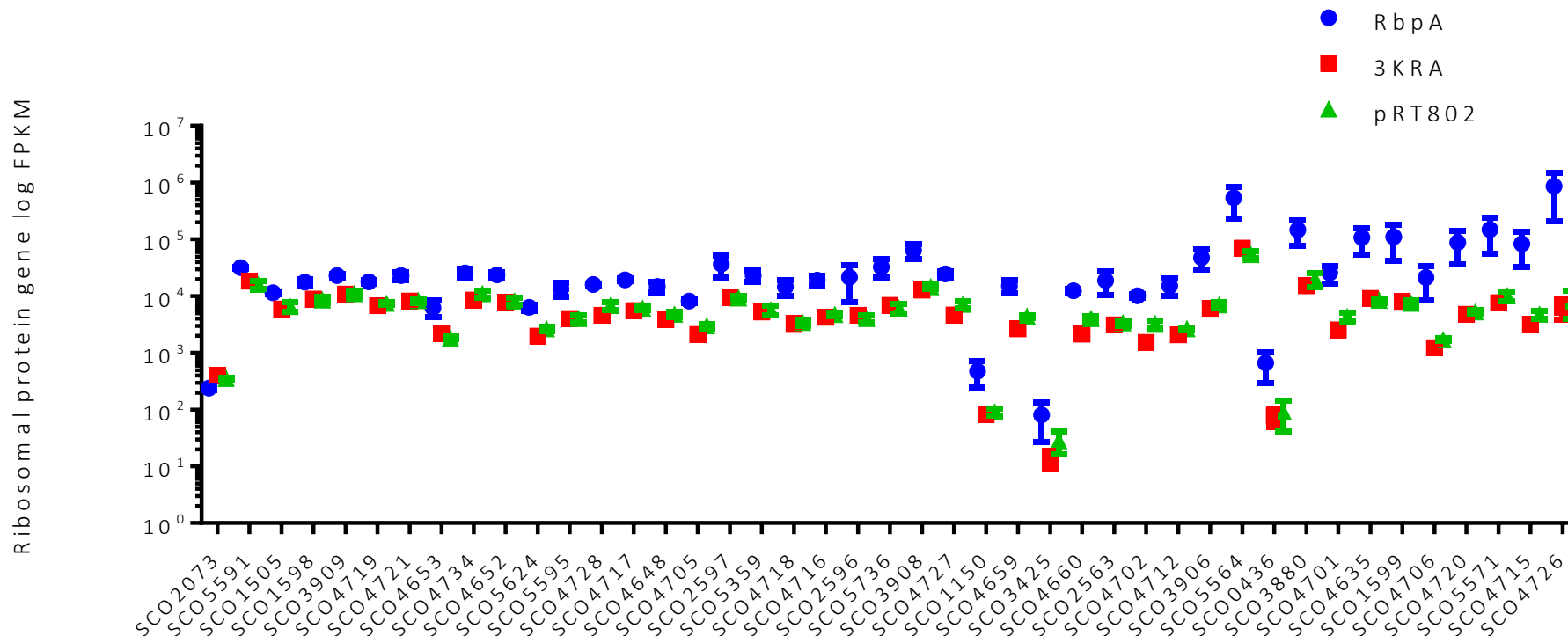


Figure 3.15. Down-regulation of ribosomal protein genes in S401 (pRT802::*rbpA*^{3KRA}) and S401 (pRT802). The ribosomal protein gene FPKM values were comparing S401 (pRT802::*rbpA*^{WT}) vs S401 (pRT802::*rbpA*^{3KRA}) vs S401 (pRT802). This data is compiled from three biological replicates and the error bars represent mean with standard deviation.

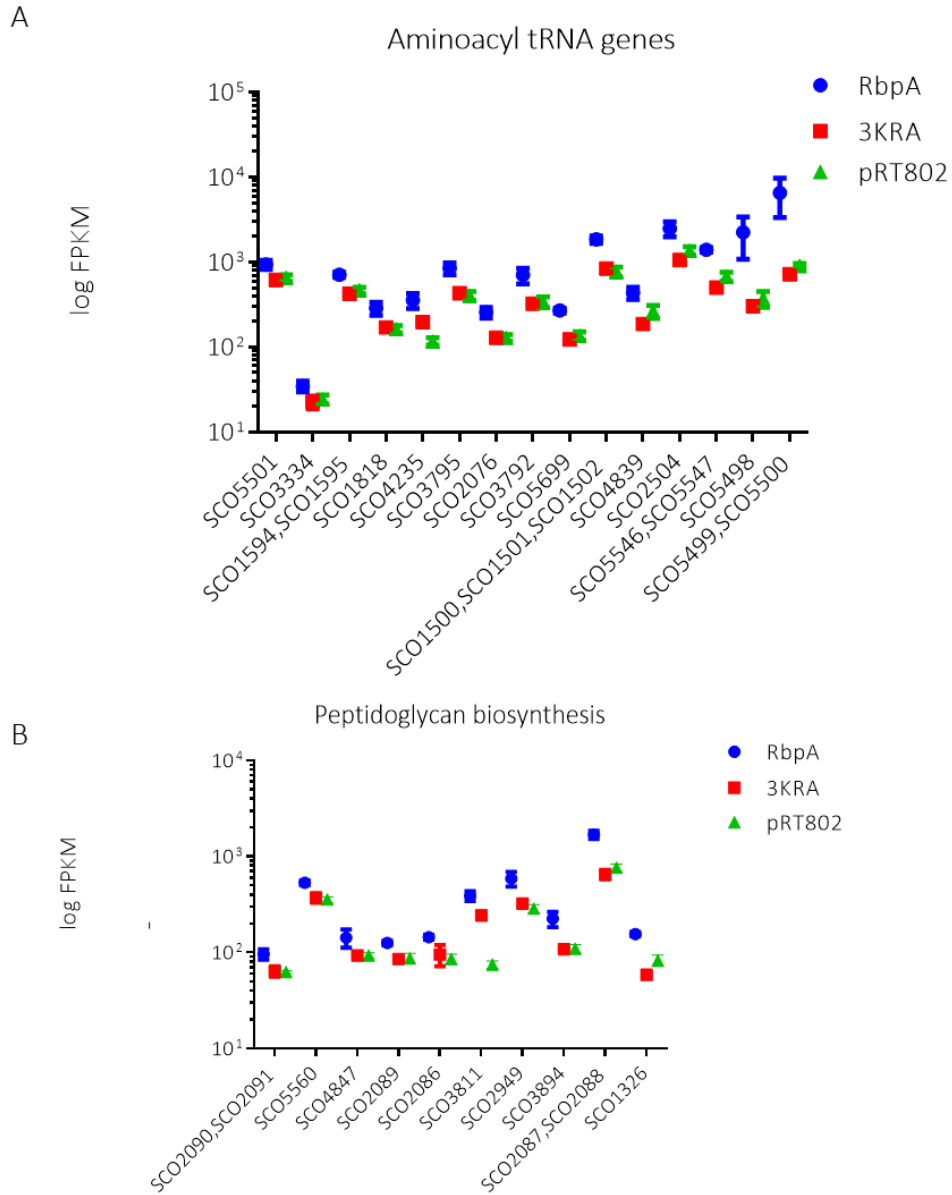


Figure 3.16. Growth-related genes are down-regulated in S401 (pRT802::*rbpA*^{3KRA}) and S401 (pRT802). (A) Aminoacyl tRNA biosynthesis genes and (B) peptidoglycan biosynthesis gene log FPKM values comparing S401 (pRT802::*rbpA*^{WT}; blue circles) to S401 (pRT802::*rbpA*^{3KRA}; red squares) and S401 (pRT802; green triangles). The results are compiled from three biological replicates and error bars represent mean with standard deviation.

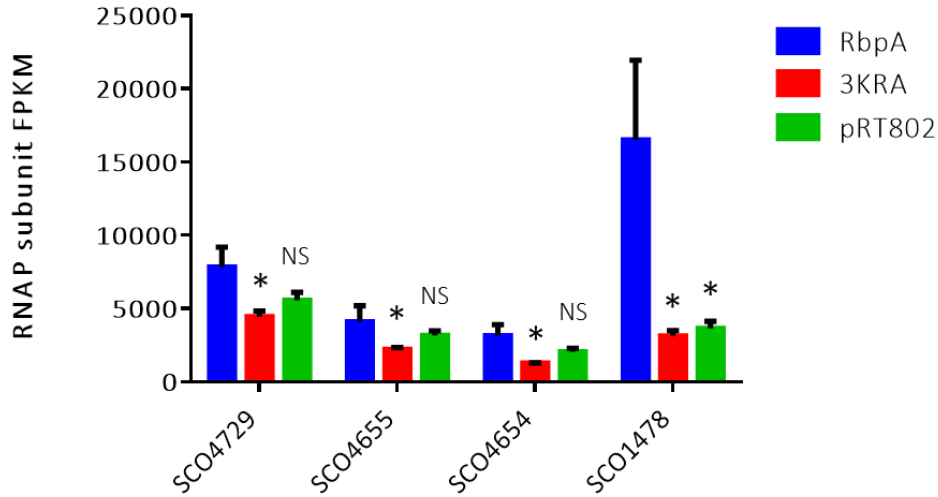


Figure 3.17. The down-regulation of RNAP subunits in S401 (pRT802::*rbpA*^{3KRA}). FPKM values comparing S401 (pRT802::*rbpA*^{WT}) to S401 (pRT802::*rbpA*^{3KRA}). The results are compiled from three biological replicates and the error bars represent mean with standard deviation. SCO4729 = α subunit, SCO4655 = β' subunit, SCO4654 = β subunit and SCO1478 = ω subunit.

The number of ribosomal protein genes that are down-regulated in the S401 (pRT802::*rbpA*^{3KRA}) and S401 (pRT802) suggested that they may be dependent on RbpA. Therefore, the promoters for the 26 available TSS for the ribosomal protein genes were aligned in order to identify features of the promoter which could indicate possible dependency on RbpA (Jeong et al., 2016). The alignment of the ribosomal protein gene promoters showed a good -10 consensus (TAnnnT) sequences for the majority, whilst the -35-promoter element is not identifiable (Fig. 3.18A). The promoter elements were compared to ribosomal genes which were unchanged, only two of the four have TSS annotated (Fig. 3.18A), SCO4735 and SCO1998. Additionally, the ribosomal gene promoters significantly down-regulated in S401 (pRT802::*rbpA*^{3KRA}) consist of multiple thymine residues upstream from the -35 element which is important for σ_2 binding.

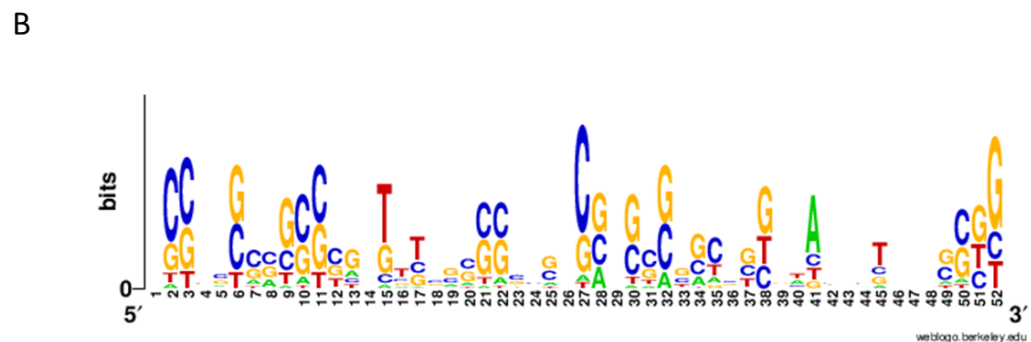
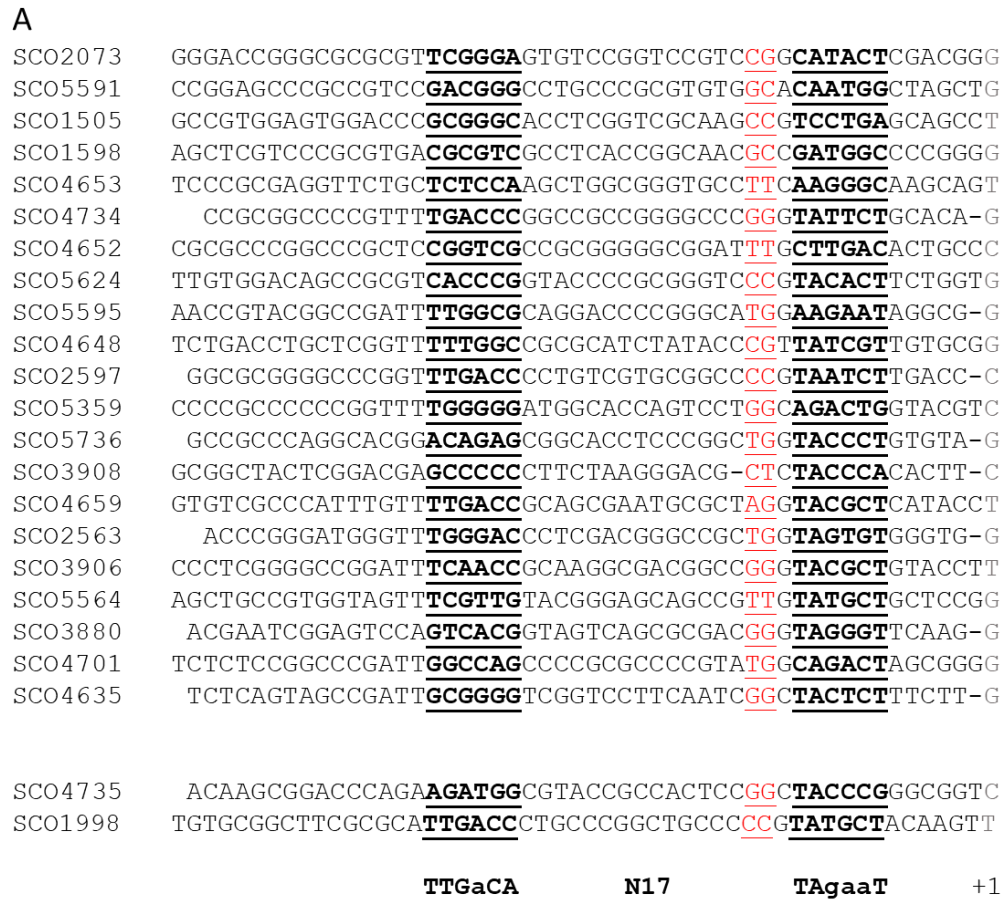


Figure 3.18. Alignment of the 22 ribosomal protein gene promoters downregulated in S401 (pRT802::*rbpA*^{3KRA}) and S401 (pRT802). The TSS were identified using supplemental data, datasheet S2 (Jeong et al., 2016). (A) Position +1 from TSS is shown in grey, -10 and -35 promoter elements are underlined, the extended -10 elements are shown in red. The consensus -10 and -35 promoter elements are shown in bold and have been aligned based on the -10 and -35 HrdB dependant conservation (Kang et al., 1997). (B) Motif conservation created using WEBLOGO (version 2.8.2), position 1 represents -52 and position 52 at the 3' presents +1 of the promoter.

The TCA cycle also referred to as the 'Krebs cycle' is used by aerobic organisms to generate energy. Thus far, most of the growth related genes are downregulated in S401 (pRT802::*rbpA*^{3KRA}) and S401 (pRT802), which could be a result of reduced ATP generation from the TCA cycle. Therefore, the RNA-seq data was used to group the TCA cycle genes together, which confirmed that 27 (out of 48) are all downregulated in S401 (pRT802::*rbpA*^{3KRA}). The S401 (pRT802::*rbpA*^{3KRA}) was compared to S401 pRT802, which matched the downregulation of 20 (out of the 27) TCA cycle genes (Fig. 3.19).

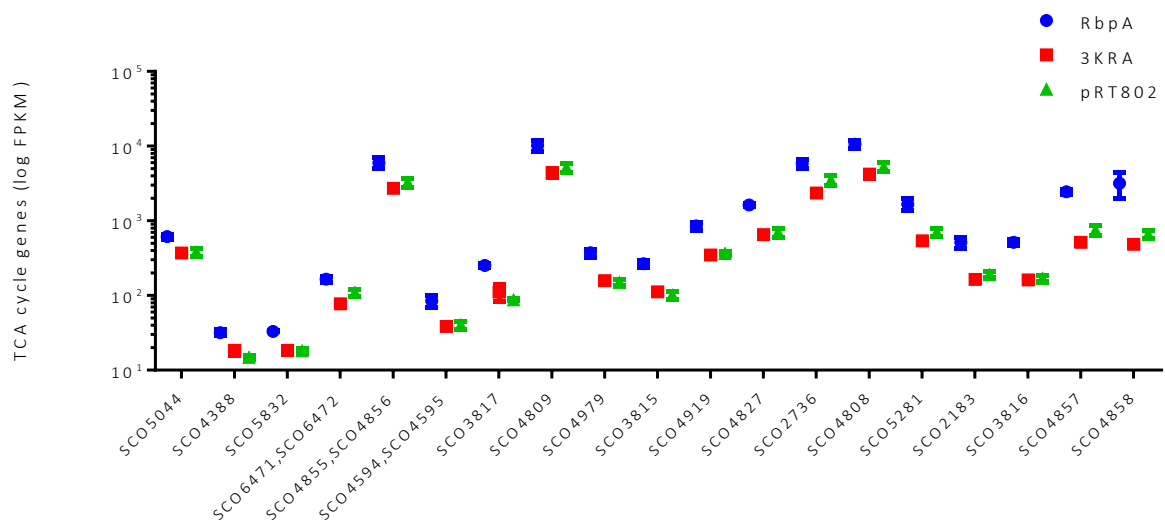


Figure 3.19. The down-regulation of the TCA cycle genes in S401 (pRT802::*rbpA*^{3KRA}) and S401 (pRT802). The FPKM values are comparing S401 (pRT802::*rbpA*^{WT}) to S401 (pRT802::*rbpA*^{3KRA}) and S401 (pRT802). The results are compiled from three biological replicates and the error bars represent mean with standard deviation.

Ribonucleotide reductase (RNR) are involved in the biosynthesis of deoxyribonucleotides, the precursor for DNA synthesis (Reichard & Ehrenberg., 1983). Class Ia RNRs consist of two homodimeric proteins, $\alpha_2\beta_2$, the large α subunit is encoded by *nrdA* and the β subunit is encoded by *nrdB*. Class II uses vitamin B12 and generally operates in an oxygen-independent process. The *Streptomyces* spp. can encode for both class I and class II RNRs and are oxygen dependant or independent, this section focuses on the class I RNR (Borovok et al., 2002). The *S. coelicolor*, *nrdA* (SCO5226) and *nrdB* (SCO5225) are part of the class Ia RNR and are oxygen dependent. The highest oxygen concentrations are present during early exponential phase compared to mid or late exponential phase (Borovok et al., 2002). Immediately upstream of

nrdA is an ORF encoding a 91-amino acid protein, *nrdX* (SCO5227). The *nrdA*, *nrdB* and *nrdX* operon is down regulated in S401 (pRT802::*rbpA*^{3KRA}) and S401 (pRT802), Fig.3.20 demonstrates the reduction in FPKM value for both RNA-seq datasets. The *S. coelicolor nrdJ* belongs to the class II oxygen independent ribonucleotide reductase (Borovok et al., 2002). *nrdJ* mRNA levels generally remain constant over the entire exponential phase and may play an important role during vegetative growth. The *nrdJ* *S. coelicolor* gene is co-transcribed in an operon with *nrdR* and is vitamin B12 dependant, $\Delta nrdJ$ fails to grow in the presence of vitamin B12. The RNA-seq data shows down-regulation of *nrdJ* (SCO5805) in S401 (pRT802::*rbpA*^{3KRA}) and S401 (pRT802) strain by 2.4 and 1.9-fold respectively (Fig. 3.21).

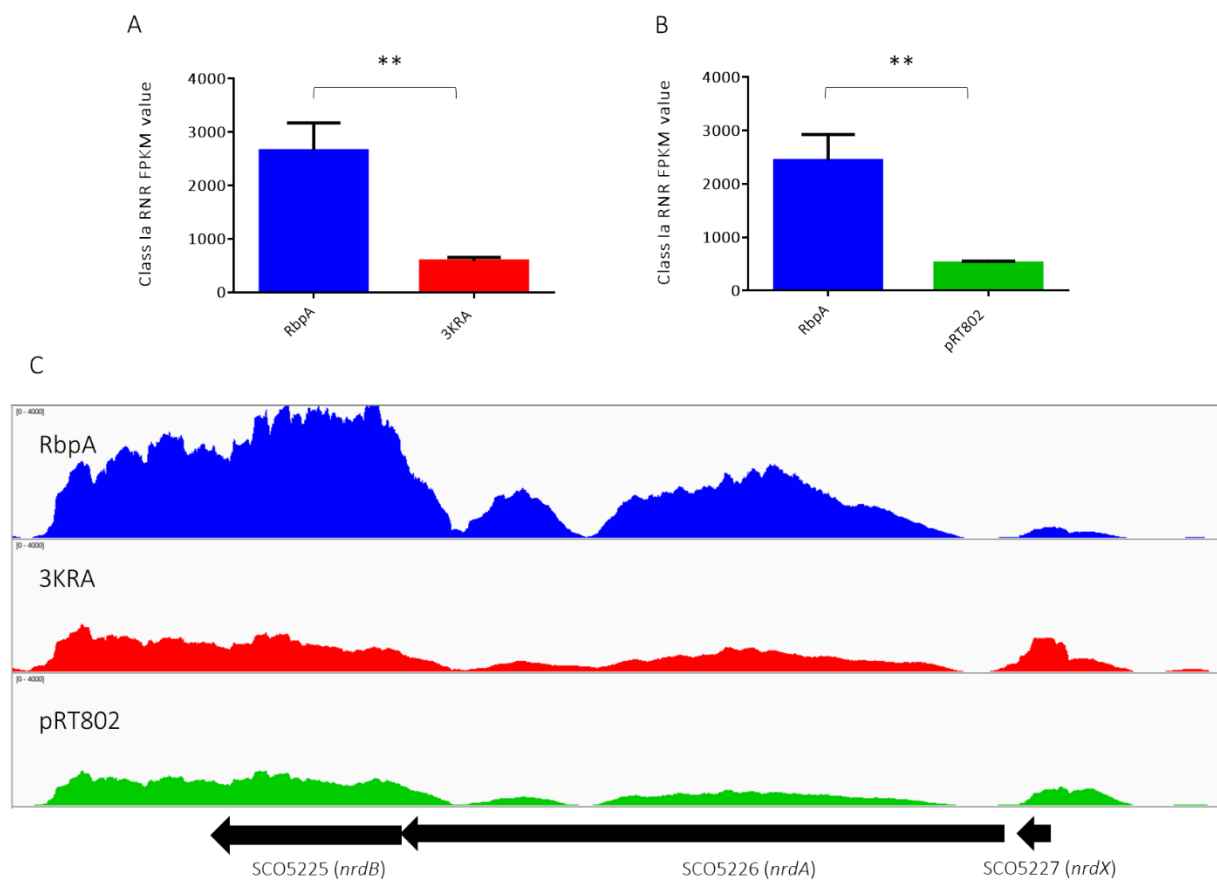


Figure 3.20. Down-regulation of class Ia RNR, *nrdABX* operon. (A) The FPKM values of *rbpA*^{WT} and *rbpA*^{3KRA}, (p-value = 0.0009). (B) The FPKM values of *rbpA*^{WT} and pRT802, (p-value = 0.0004). This data is compiled from three biological replicates and the error bars denote mean with standard deviation.

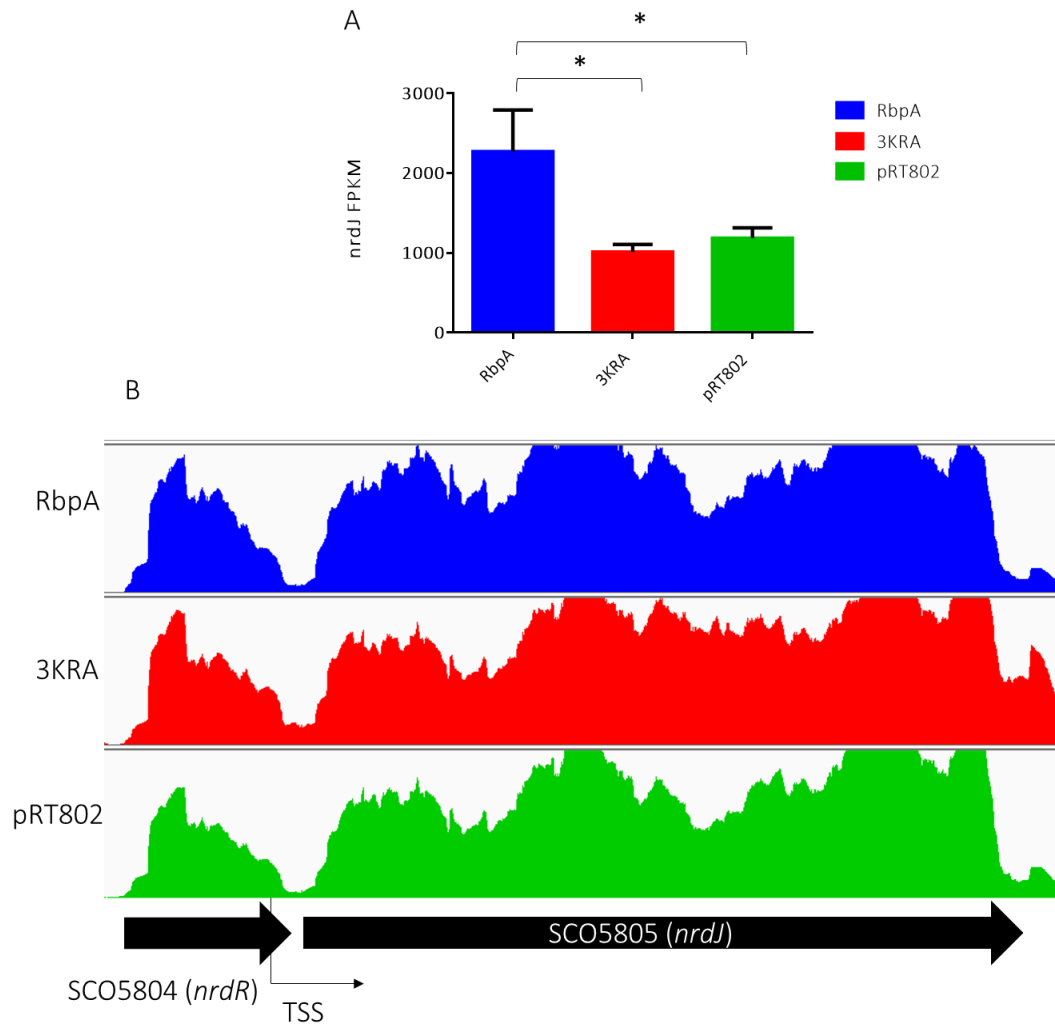


Figure 3.21. Vitamin B12 dependant ribonucleotide reductase, *nrdJ* is down-regulated in S401 (pRT802::*rbpA*^{3KRA}) and S401 (pRT802). (A) Visualisation of the RNA-seq reads at the SCO5805 location using IGV (version 2.4.9) with TSS of SCO5805 shown using TSS data from Jeong et al (2016) (B) FPKM values for *rbpA*^{WT} and pRT802, this data is compiled from three biological replicates, *RbpA*^{WT} vs *RbpA*^{3KRA} =p-value 0.015 and *RbpA*^{WT} vs pRT802 =p-value 0.0255.

3.1 Discussion

3.2.1 The RbpA BL does not have an essential function

RbpA together with RbpB confers an essential function in *S. coelicolor*. The RbpA BL is highly conserved across the Actinobacteria and therefore the first key aim of this section was to determine whether the BL-DNA interactions are the essential role of RbpA. The BL residues are required *in vitro* for transcription activation, but it is not clear whether they are essential *in vivo*. The RbpA^{R79A} failed to activate transcription *in vitro* and produced a small colony phenotype *in vivo* (Hubin et al., 2015; 2017). The *S. coelicolor* RbpA^{R80} mutant reduces growth compared to the RbpA^{WT} but the growth phenotype does not appear to be as severe as a $\Delta rbpA$ null mutant. This could be because the lysines contribute to RbpA function, perhaps playing an overlapping role when R80 is deleted. Additionally, a crystal structure of RbpA/TIC in *M. smegmatis* has shown that the highly conserved RbpA^{R79} makes polar interactions with negatively charged -13 and -14 bases of the non-template promoter DNA (Hubin et al., 2017a). The highly conserved RbpA^{K76} (K77, *S. coelicolor*) makes long electrostatic interactions with -17/-18 of the template strand (Hubin et al., 2017a). This provides evidence of a potential role of the RbpA BL in promoter binding (Hubin et al., 2017a; Prusa et al., 2018). Therefore, the RbpA^{3KRA} mutant was engineered which substitutes all three lysines (K74, K75 and K77) and R80 with alanine. The mutants were tested in the conditional mutant S401 which, would reveal the activity of the mutants without the complication of *rbpB*. The S401 (pRT802::*rbpA*^{3KRA}) demonstrated a small colony phenotype with ACT production compared to S401 (pRT802::*rbpA*^{WT}) but again the phenotype was not as severe as a $\Delta rbpA$ null mutant. This shows that the BL, and by extension BL-DNA interactions, do not constitute the essential function of RbpA. Importantly, this result eludes to additional roles that are independent of the essentiality of RbpA. Next, the RbpA BL strains were cultured in liquid media to identify an effect on growth. The S401 (pRT802::*rbpA*^{3KA}) grew at a similar rate to S401 (pRT802::*rbpA*^{WT}) whilst S401 (pRT802::*rbpA*^{R80A}) demonstrated some reduction. Most interestingly, S401 (pRT802::*rbpA*^{3KRA}) demonstrates delayed entry into log phase and reduced growth rate overall compared to S401 (pRT802::*rbpA*^{WT}). This provided further evidence of a potential combined function of the three lysine residues and R80.

Additionally, during the period of this study several studies were published providing an insight into the possible role of the RbpA BL. The BL residues were annotated onto a *M. tuberculosis* RbpA-SID RPo structural model and provided the first evidence of the RbpA BL residue interactions with DNA (Hubin et al., 2015). This was later followed up by a RbpA-transcription initiation X-ray crystal structure (Hubin et al., 2017) and subsequently the cryo-EM RNAP- σ^A -RbpA-Fdx structure confirmed that the RbpA BL residues are positioned closely to the DNA phosphate backbone upstream of the -10 element (Boyaci et al., 2018). In vivo studies showed that RbpA^{R79A} is not morphologically different to RbpA^{WT} despite the slow growth phenotype. Finally, kinetic studies show that RbpA stabilises RPo complexes (Rammohan et al., 2015) and the RbpA^{R79A} delays transition of RPo to RPo (Hubin et al., 2017), and is unable to stabilise the RPo complex (Prusa et al., 2018). Most interestingly, the cryo-EM structure showed that the RbpA-NTT threads into the RNAP active site and interacts with conserved residues of the σ finger ($\sigma_{3.2}$ linker; Boyaci et al., 2018).

3.2.2 Increased susceptibility of RbpA BL mutants to rifampicin treatment

The *S. coelicolor* $\Delta rbpA$ mutant S101 exhibits a 15-fold increase in susceptibility to rifampicin compared to *rbpA*^{WT}, although over-expression of RbpA did not further increase rifampicin resistance (Newell et al., 2006). Therefore, this study investigated the contribution of the RbpA BL to the susceptibility to rifampicin compared to the previous study on the $\Delta rbpA$ mutant. In support of the previous study, S401 pRT802 did show increased susceptibility to rifampicin compared to S401 (pRT802::*rbpA*^{WT}). The S401 (pRT802::*rbpA*^{R80A}) and S401 (pRT802::*rbpA*^{3KRA}) also show increased susceptibility but not as severe as S401 (pRT802). The RbpA NTT does thread through to the active site of RNAP, therefore direct contact between RbpA and rifampicin cannot be ruled out (Boyaci et al., 2018). Alternatively, RbpA may bind more tightly to RNAP through conformational changes in the presence of rifampicin, however abortive initiation assays prove that RbpA does not affect the binding of rifampicin to RNAP (Hu et al., 2014). We know that, rifampicin causes the activation of SigR, thus causing the activation of SigR regulon. RbpA consists of a *sigR*-dependant promoter which results in the up-regulation of RbpA, which would ensure survival of the cells in response to rifampicin. The increased susceptibility of $\Delta rbpA$ to rifampicin remains unclear. One possible solution for this is to construct a *sigR* mutant or alternatively to delete the *sigR* dependant *rbpA* promoter.

Fidaxomicin (Fdx) is an RNAP inhibitor used clinically for *C. difficile* infection and has now been shown to have inhibitory effects against *M. tuberculosis in vitro* (Boyaci et al., 2018). Cryo-EM structures show RNAP-RbpA interacting with Fdx. Fdx acts like a doorstop which jams RNAP in an open complex, preventing promoter DNA from entering the RNAP active site (Boyaci et al., 2018). The Fdx binding site is independent of rifampicin and there is no cross-resistance with rifampicin. Resistance to Fdx increases 35-fold when the RbpA-NTT is truncated (Boyaci et al., 2018) and therefore it is quite plausible that RbpA-NTT occupying the RNAP active-site, prevents binding of rifampicin to the active site thus resulting in resistance. Although, the RbpA-NTT is not required for viability, changes in the other RbpA domains may affect the positioning of the NTT within the RNAP active site. Similarly, a recent publication characterises a regulatory pathway for isoniazid tolerance in *M. smegmatis* (Wang et al., 2018). RbpA binds to the stress induced group II sigma factor, σ^B and in turn activates the transcription of a polyphosphate kinase gene, *ppk1*. This resulted in an increase in intracellular polyphosphates and drug tolerance to isoniazid (Wang et al., 2018). The problem with such studies is that RbpA is an essential protein and thus has pleiotropic effects on the cell, it is therefore no surprise that this pathway has been identified but does not prove that RbpA is directly responsible for isoniazid tolerance.

3.2.3 The RbpA BL mutant and $\Delta rbpA$ cause down-regulation of growth related genes

This study has found that S401 (pRT802::*rbpA*^{3KRA}) and S401 (pRT802) cause down-regulation of growth related genes, these include ribosomal protein genes, TCA cycle genes, aminoacyl-tRNA biosynthesis and peptidoglycan biosynthesis genes. This could be because S401 (pRT802::*rbpA*^{3KRA}) and S401 (pRT802) fail to grow at the same rate as S401 (pRT802::*rbpA*^{WT}), thus failure to grow causes global transcriptional down-regulation. Alternatively, the loss of the essential gene *rbpA* could be responsible for the global down-regulation of growth related genes. Upon closer inspection of the ribosomal protein genes, genes that were down-regulated mostly lack a -35-consensus sequence, which suggests that RbpA may be important for promoters that lack a consensus -35 sequence.

Ribonucleotide reductases (RNRs) which catalyse the conversion of ribonucleotides to deoxyribonucleotides are also down-regulated in S401 (pRT802::*rbpA*^{3KRA}) and S401 (pRT802). The oxygen dependent ribonucleotide reductases may contribute to the slow growth phenotype of S401 (pRT802::*rbpA*^{3KRA}) and S401 (pRT802). The global ribonucleotide reductase regulator, *nrdR* can bind to the promoters of both class Ia and II RNRs (Borovok et al., 2006). In addition to this, the B12 riboswitch provides regulation of the class Ia RNR and serves as a cofactor for the class II RNRs (Borovok et al., 2002; 2006). Therefore, in summary it is difficult to comment on whether the RbpA BL and Δ *rbpA* cause the changes in the transcriptome which lead to down-regulation of growth-related genes or whether slow growth causes the transcriptomic changes. Finally, the results of the RNA-seq indicate that the RbpA BL is required for efficient gene expression.

3.2.4 Future work

Firstly, the potential RbpA^{3KRA} targets identified from the RNA-seq have not been confirmed with qRT-PCR yet. The ribosomal protein genes will also be confirmed using *in vitro* transcription assays to investigate the dependency of promoters on RbpA^{WT} and RbpA^{3KRA}. The optimisation and purification of core RNAP and σ^{HrdB} has been completed, therefore RbpA^{WT} and RbpA^{3KRA} remains to be purified (Appendix. 2). Secondly, to establish a link between the genes involved in ACT production, the S401 pRT802::*rbpA*^{3KRA} and S401 pRT802 strains need to be cultured for 72-120 h before RNA is extracted for qRT-PCR examination of *actII-ORF4*, *adpA*, *dasR*, *absA2* and *atrA*. Lastly, this study has provided preliminary evidence that RbpA may compensate for the loss of a consensus -35 promoter sequence, therefore these need further validation. The RbpA^{DAS+4} and CarD^{DAS+4} depletion system (Results chapter IV) could be implemented in this study by incorporating promoters with a 'good' or 'poor' -35 into a GFP expression plasmid to test the level of expression of the promoters using S1 nuclease protection assays.

Results chapter II: Investigating the
role of the RbpA basic linker in
Mycobacterium tuberculosis

4.0 Overview of the RbpA basic linker in *M. tuberculosis*

RbpA (Rv2050) is an essential protein in *M. tuberculosis* and shares 56% end-to-end amino acid similarity with the *S. coelicolor* RbpA (Bortoluzzi et al., 2013; Sassetti et al., 2003). Structural studies suggest that the key role of RbpA is to contact DNA upstream of the -10 elements, stabilising promoter open complexes (Hu et al., 2012). This is hypothesised to be critical for the action of some promoters and therefore disruption of this essential function affects mycobacterial viability. However, mutagenesis of the RbpA basic linker residues in *S. coelicolor* demonstrated that removal of this DNA binding interaction does not impact viability, suggesting that this function may be compensated for or that RbpA plays a different role. RbpA in *S. coelicolor* exists in tandem with its paralogue, RbpB thus far no such paralogue has been found in *M. tuberculosis*. Since RbpA is an essential gene in *M. tuberculosis*, we investigated the role of RbpA in transcription initiation in *M. tuberculosis* by constructing strains in which the only expressed copy of *rbpA* contained a series of mutations of the basic linker. This study identified that the *M. tuberculosis* RbpA^{3KRA} was unable to grow on solid media compared to wild type or other basic linker mutants, however the growth defect was not observed in liquid culture. This differential culturability has been associated with persister phenotypes in *M. tuberculosis* (Mukamolova et al., 2010; Salina et al., 2014). To investigate this phenotype further and to understand the impact of modifying the RbpA basic linker on the *M. tuberculosis* transcriptome, we compared RbpA^{3KRA} to RbpA^{WT} using RNA-sequencing. This chapter therefore focuses on addressing the biological importance of the RbpA basic linker (BL) in *M. tuberculosis*.

4.1.0 Construction of RbpA BL mutants in *M. tuberculosis*

To identify essential genes in *M. tuberculosis* a study constructed a transposon library using the *Streptomyces* Pristinamycin I-inducible *ptr* promoter in a Tn-*pip/pptr* system. Pristinamycin inhibits ribosomes and the pristinamycin resistance gene promoter, *ptr* encodes for a multidrug efflux pump (Ehrt & Schnappinger, 2014). The *ptr* promoter is repressed by the transcription factor Pip and can be activated by pristinamycin (Ehrt & Schnappinger, 2014). The transposons were inserted into the mycobacteriophage phAE87, transduced into *M. tuberculosis*, which created a transposon insertion library (Forti et al., 2011). This screening

confirmed the essentiality of *rbpA* in *M. tuberculosis* with the insertion of the Tn-*pip/pptr* system 45 bp upstream of the *rbpA* open reading frame by sequencing (Forti et al., 2011; kindly gifted by Dr Helen O'Hare, University of Leicester). The *rbpA* transposon mutant is stabilised with hygromycin and pristinamycin (PI) to maintain native *rbpA* expression, this strain will be referred to as M233. The objective was to express a series of RbpA basic linker mutant alleles in this PI-inducible system by integrating these mutated RbpA alleles into the genome at the mycobacteriophage L5 *attB* site using the integrative *E. coli-Mycobacterium* shuttle plasmid pMV306. The addition of PI causes expression of the inducible wild type chromosomal copy of *rbpA*, however when PI is removed only the integrative *rbpA* allele (from pMV306) should be expressed. This allowed investigation of the functional significance of the RbpA BL mutants in an essential *M. tuberculosis* gene. The coding region of *rbpA* together with the 357 bp upstream promoter region (743 bp in total) was cloned into EcoRV-digested pBluescript II SK+ and confirmed by sequencing. The resulting plasmid, pMV306::*rbpA*^{WT} was used as a template for inverse PCR mutagenesis to create the BL mutations 3KA (K74, K75 and K77 substituted with alanine), R79A (R79 substituted with alanine) and 3KRA (K74, K75, K77 and R79 substituted with alanine). The inverse PCR mutagenesis primers are listed in Table 2.5. Following DNA sequence analysis confirming the expected nucleotide sequences, the three *rbpA*-mutant sequences along with wild-type *rbpA* and empty vector (containing no *rbpA* sequence) were cloned as EcoRI-HindIII fragments into pMV306. The constructs were confirmed by sequencing and electroporated into *M. tuberculosis* M233 (Tn-*pip/pptr* Δ *rbpA*). The recombinants were plated onto Middlebrook 7H10 agar supplemented with OADC (Oleic acid-Albumin-Dextrose-catalase), kanamycin, pristinamycin I and hygromycin. Insertion of the mutant alleles at the *attB* site were confirmed by colony PCR (Methods; 2.3.2) using primers MTB2_F and MTB2_R (Table 2.5) that amplify 8 bps from pMV306 into the ORF of RbpA of the integrated pMV306::*rbpA*. The recombinant strains were cultured for 7-days with kanamycin, hygromycin and PI (to express the inducible chromosomal wild type *rbpA* allele that is essential for growth). Thus, in the presence of PI, Pip is no longer able to repress *pptr* allowing expression of the native WT *rbpA*. In the absence of PI, *pip* is expressed from the *pfurA102* promoter that represses *ptr* inhibiting the expression of that native WT *rbpA* (Fig.4.1). This scenario allowed the impact of the integrated pMV306::*rbpA*^{WT}, pMV306::*rbpA*^{3KA}, pMV306::*rbpA*^{R80A}, and pMV306::*rbpA*^{3KRA} to be assessed.

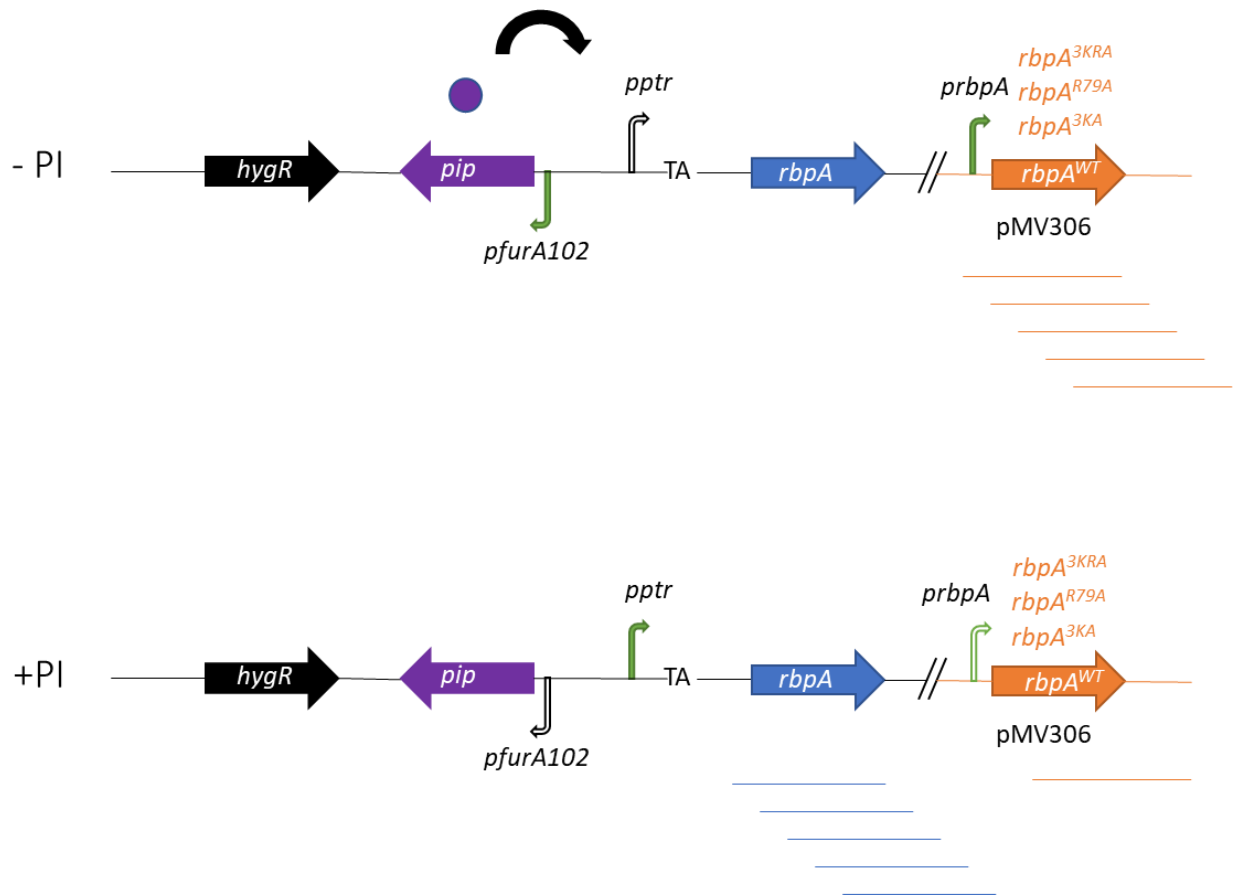


Figure 4.1. The regulatory control of the *rbpA* conditional mutant in *M. tuberculosis*. The transposon *rbpA* conditional mutant strain was created by Francesca Forti (Forti et al., 2011), the pMYT829 construct consists of the *hyg* resistance cassette (black, marked *hyg*), *pip* (*ptr* repressor, in purple) controlled by the promoter *pfurA102*, *pptr*, hyperactive transposase and the left and right inverted repeats of the Himar1 transposon. The transposon has inserted at the TA site 43 bases upstream of the translational start site of *rbpA*. The *pfurA102* promoter (arrow with black outline) and *pip* (purple) are integrated at the Φ L5 *att* site on the *M. tuberculosis* chromosome. pMV306 (orange) is an integration plasmid, carrying kanamycin cassette and recombinant *rbpA* mutant alleles integrated at the *attP* site. In the absence of PI, the only *rbpA* copy expressed is the pMV306 integrated copy (green arrow represents the activated *pfurA102* and *prbpA* promoters, *pip* is expressed which represses the *rbpA* promoter, *pptr*. The presence of PI prevents *pip* expression from the *pfurA102* promoter and allows expression of native PI dependent WT *rbpA* expression (green arrow represents the activated *pptr* promoter).

4.1.1 Growth curves of the BL mutants in *M. tuberculosis*

The RbpA BL mutants in the absence of RbpB in *S. coelicolor* exhibited a growth phenotype displayed by small colony formation and ACT production, therefore growth curves were performed for the RbpA BL mutants in *M. tuberculosis*. To investigate the impact of the RbpA BL mutants on *M. tuberculosis* growth, optical density and CFUs (colony forming units) were measured over 28 days. The strains were cultured for seven days in the presence of PI (allowing WT *rbpA* expression), then washed with PBS to remove all PI. The cells from the RbpA BL strains were normalised using optical density to ensure that each recombinant growth curve was started with the same number of cells. Cultures were plated onto LB agar and inoculated into liquid LB media to check for contamination; the results of which confirmed no contamination. The stability of the inserted transposon and the absence of point mutations in the pristinamycin I gene in the engineered strains was checked by designing primers 52 bp upstream of the TA insertion site (PTRF) and the reverse primer at the C-terminus of native *rbpA* (Table 2.5). These primers amplified the *pptr*, transposon insertion site and the chromosomal copy of *rbpA* to check for spontaneous mutations. The sequencing results confirmed the presence of the transposon insertion of the Tn-pip-pptr system and showed no SNPs. The integrated pMV306::*rbpA*^{WT} and BL mutants were also PCR-amplified and sequenced, using primers MTB2_F and MTB2_R (Table 2.5). Similarly, the sequencing results confirmed that pMV306::*rbpA*^{WT} contained RbpA^{WT} and the mutants contained the relevant mutant alleles. Samples were taken through the growth curves at regular intervals to measure mycobacterial growth by OD₆₀₀ (Fig. 4.2). The RbpA BL mutants did not show any growth defect compared to RbpA^{WT} as measured by optical density over the 28-day period (Fig. 4.2). Unexpectedly, the empty vector control, M233 (pMV306) containing no recombinant *rbpA* grew at a similar rate to M233 (pMV306::*rbpA*^{WT}). This is likely a result of leaky *rbpA*^{WT} expression from the conditional system in the absence of PI.

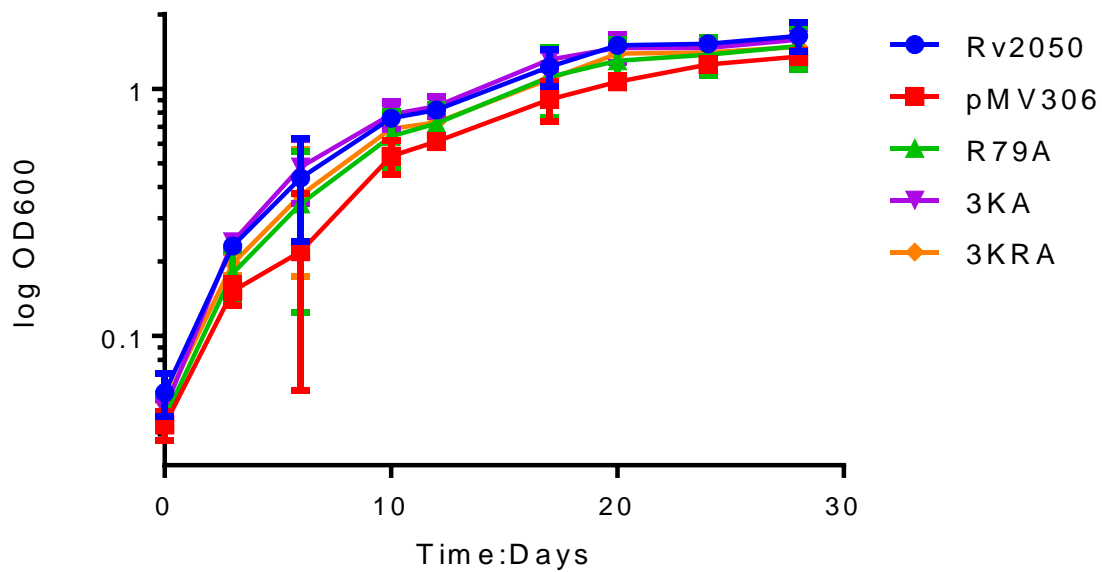


Figure 4.2. Growth of the *rbpA* BL mutants alongside *rbpA*^{WT} and empty vector in *M. tuberculosis* in the absence of PI, measured by optical density. M233 (pMV306::*rbpA*^{WT}; blue), M233 (pMV306- empty vector; red), M233 (pMV306::*rbpA*^{R79A}; green), M233 (pMV306::*rbpA*^{3KA}; purple) and M233 (pMV306::*rbpA*^{3KRA}; orange) cultured in 7H9 (supplemented with ADC) at 37°C for 28 days. This figure is compiled from three biological replicates and the error bars represent mean with standard deviation.

4.1.2 Confirmation of complementation of the BL strains in *M. tuberculosis*

The growth curve (measured by optical density) did not show a phenotype for the RbpA null empty vector control, which raised the question of whether the PI inducible system was operating effectively. Therefore, complementation of the RbpA BL strains, \pm PI, was tested on solid agar. The BL strains were cultured for 7 days (OD₆₀₀ ~0.3) and then the cells were washed and normalised followed by serially diluting from neat to 10⁻³ and plated onto 7H10 in the absence/presence of PI (Fig. 4.3). M233 (pMV306- empty vector), M233 (pMV306::*rbpA*^{3KA}), M233 (pMV306::*rbpA*^{R79A}) and M233 (pMV306::*rbpA*^{3KRA}) all cultured the same as M233 (pMV306::*rbpA*^{WT}) in the presence of PI, demonstrating that the PI inducible system is operational from the chromosomal *rbpA*^{WT} location. In the absence of PI, the M233 (pMV306) strain lacking a recombinant *rbpA* did not grow, showing that there was no expression from the *pptr* inducible *rbpA*^{WT} allele in the absence of PI. This was expected as *rbpA* is an essential

gene in *M. tuberculosis*. The M233 (pMV306::*rbpA*^{WT}) grew in a similar fashion to when PI was present, demonstrating that, expression from the integrated pMV306::*rbpA* successfully complemented in the absence of PI (Fig. 4.3). M233 (pMV306::*rbpA*^{3KA}) grew similarly to M233 (pMV306::*rbpA*^{WT}), on solid media. This correlates with the phenotypic data of the same RbpA^{3KA} mutations in *S. coelicolor*. M233 (pMV306::*rbpA*^{R79A}) also grew in a similar fashion to M233 (pMV306::*rbpA*^{WT}), however M233 (pMV306::*rbpA*^{3KRA}) demonstrated a difference in growth at dilutions 10⁻² and 10⁻³ compared to M233 (pMV306::*rbpA*^{WT}). These experiments measuring growth on solid media suggested that *rbpA*^{3KRA} may be important for growth in *M. tuberculosis* H37Rv. M233 (pMV306::*rbpA*^{3KRA}) however was not essential for viability in *M. tuberculosis* as there was more growth than M233 (pMV306), this strongly suggested that RbpA may have additional functions in *M. tuberculosis* that impact on survival.

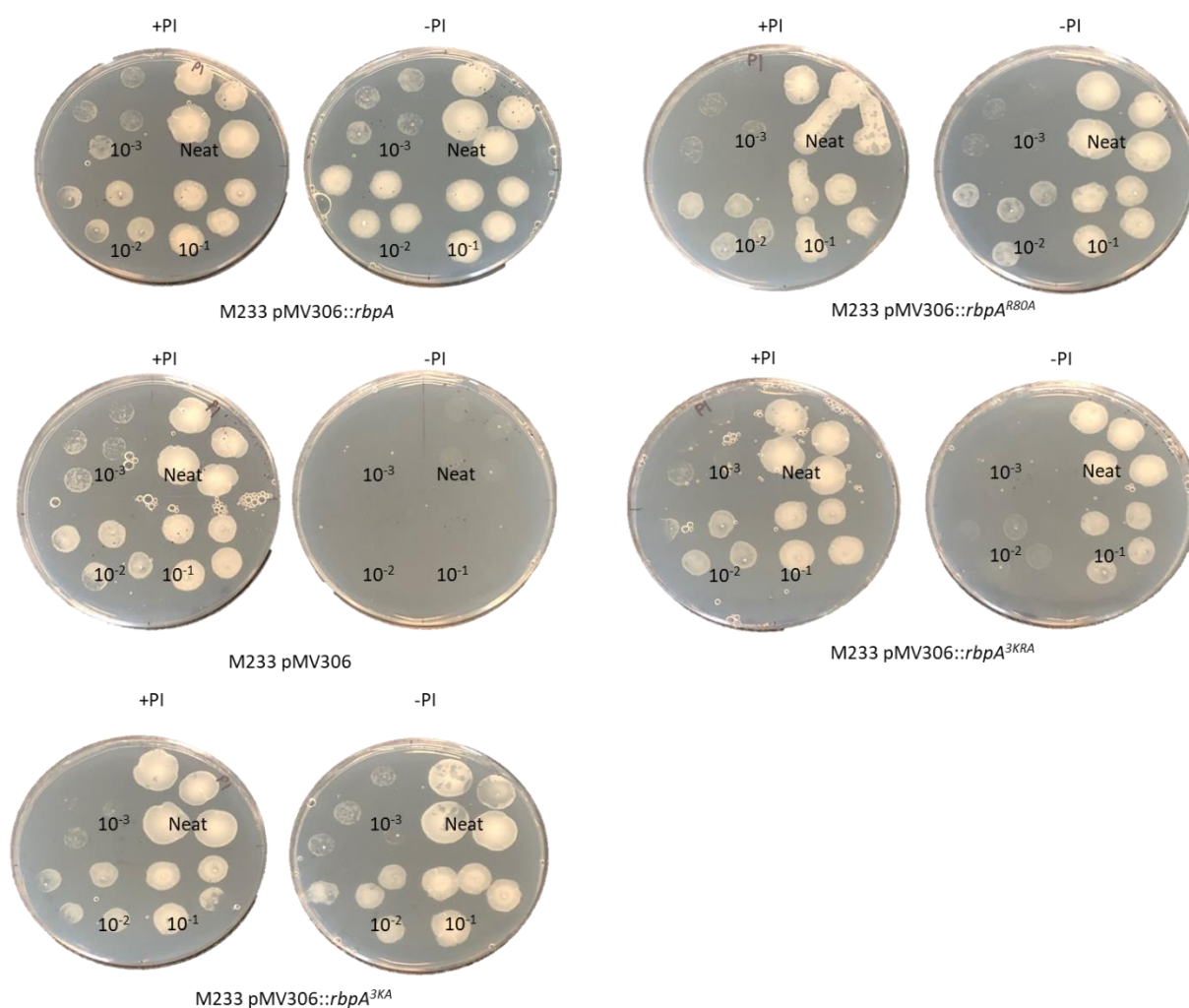


Figure 4.3. Growth of *M. tuberculosis* BL strains on solid media. M233 (pMV306::*rbpA*^{WT}; wt control), M233 (pMV306; empty vector), M233 (pMV306::*rbpA*^{3KA}), M233 (pMV306::*rbpA*^{R79A}) and M233 (pMV306::*rbpA*^{3KRA}) were plated in the presence of 0.5 µg/mL PI (left) and absence of PI (right). Dilution series; neat, 10⁻¹, 10⁻² and 10⁻³ plated onto 7H10 Middlebrook supplemented with OADC and incubated for 3 weeks at 37°C.

4.1.3 Characterising the growth phenotype of M233 (pMV306::*rbpA*^{3KRA})

The M233 (pMV306::*rbpA*^{3KRA}) strain was unable to grow on solid agar as well as M233 (pMV306::*rbpA*^{WT}) but was able to grow at a similar rate as M233 (pMV306::*rbpA*^{WT}) in liquid media (as measured by optical density). Therefore, to further explore this differential culturability, the growth curves were repeated measuring OD, CFUs and MPN (Most Probable Number) count (Fig. 4.4). MPN is a method for estimating the number of bacteria in liquid

culture by conducting a series of bacterial dilutions, incubating and then counting the number of wells that do not become turbid (Jarvis et al., 2010). Differential culturability has been associated with a persister 'non-culturable' phenotype in *M. tuberculosis* therefore the same methods were employed in this study (Mukamolova et al., 2010; Salina et al., 2014). As before no difference was observed between M233 (pMV306::*rbpA*^{WT}), M233 (pMV306) and the RbpA BL mutants in liquid media measured using optical density, as previously observed. At 0, 7, 14 and 21 days MPN and CFU counts were performed together on three biological replicates and four technical replicates for CFU counting and the same for MPN counts (Fig. 4.4). At day 21, no visible colonies on any of the agar plates were present, possibly the result of plate formulation errors, therefore the CFU and MPN results for day 21 are not shown. The CFU data revealed that M233 (pMV306) failed to grow at day 0 to 14 whilst the MPN values showed that there were $\sim 2 \times 10^7$ viable bacteria at day 7 and $\sim 2 \times 10^7$ at day 14. This suggests that the PI-inducible system repressing WT RbpA expression is only fully effective on solid media. CFU and MPN counts were correlated for M233 (pMV306::*rbpA*^{WT}) and M233 (pMV306::*rbpA*^{R79A}), at day 7 and 14. MPN was not calculated for M233 (pMV306::*rbpA*^{WT}) and M233 (pMV306::*rbpA*^{R79A}) at day 14 as both strains grew at all of the dilutions from 10^{-1} to 10^{-12} , and therefore no limiting dilution was recorded. This suggests that the R79 residue alone does not impact viability. M233 (pMV306::*rbpA*^{3KRA}) appeared attenuated at day 0 and day 7 by CFU and MPN however recovered at day 14. This could be an indication that the cells enter into a non-culturable state, or are attenuated in some way during lag phase growth after removal of WT RbpA (after resuspending in media without PI).

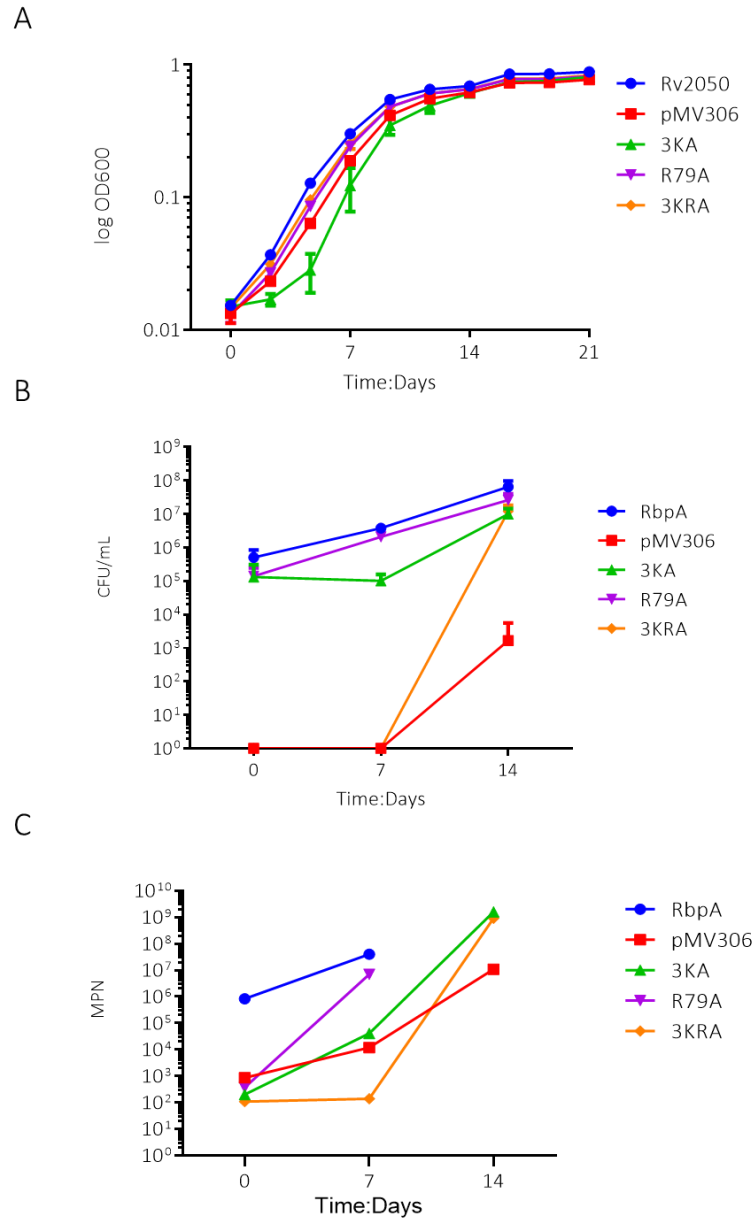


Figure 4.4. Growth rate assays for the RbpA BL mutants. (A) OD₆₀₀ readings over a 21 day period, (B) CFUs at days 0, 7 and 14, (C) MPN at days 0, 7 and 14 for M233 (pMV306::*rbpA*^{WT}; blue), M233 (pMV306; red), M233 (pMV306::*rbpA*^{3KA}; green), M233 (pMV306::*rbpA*^{R79A}; purple) and M233 (pMV306::*rbpA*^{3KRA}; orange) were cultured in the presence of 30 µg/mL kanamycin to allow expression of the integrative pMV306 plasmid and 50 µg/mL hygromycin to stabilise the transposon insertion. The OD readings are taken from three biological replicates, the CFUs and MPN readings represent three biological and four technical replicates. The results for RbpA and R79A at day 14 are not shown because all of the dilutions were turbid, absence of dilution with no growth. The CFUs for day 21 are not shown as there was no observed colonies on any agar plate. The error bars represent mean with standard deviation.

4.2.0 Profiling the impact of *rbpA*^{3KRA} substitutions on global gene expression patterns

The most striking RbpA BL mutant phenotype in *S. coelicolor* was RbpA^{3KRA} (Results Chapter I). This was replicated on solid media in *M. tuberculosis* where M233 (pMV306::*rbpA*^{3KRA}) was not viable. The incomplete PIP-mediated repression of native WT *rbpA* in liquid culture meant that this attenuated growth phenotype was not replicated in broth culture. This leaky repressor system did however present an opportunity to assay the impact of over-expressing the *rbpA*^{3KRA} variant without significantly affecting growth, unlike on solid i.e. to profile comparable samples from M233 (pMV306::*rbpA*^{WT}) and M233 (pMV306::*rbpA*^{3KRA}) in a system where complete repression would result in severe attenuation. Therefore, to explore the impact of M233 (pMV306::*rbpA*^{3KRA}) substitutions, RNA was extracted from mid-log phase M233 (pMV306::*rbpA*^{WT}) and M233 (pMV306::*rbpA*^{3KRA}) at day 8 (OD₆₀₀=0.6; Fig. 4.2; Methods: 2.6.1). RNA was quantified using the Qubit HS assay (Table. 4.1), and the quality of the RNA was assessed using the Nanodrop spectrophotometer and Agilent Bioanalyser RNA 6000 Nano kit (Fig. 4.5) before commencing RNA-sequencing. The Bioanalyser showed that the quality of the ribosomal 16s and 23s subunits were optimum and the RIN scores (>9.6) confirmed this. The Nanodrop provided an absorbance ratio of ~2 for the 260/280 and 260/230 wavelength ratios, confirming that the RNA was not contaminated with protein or salt.

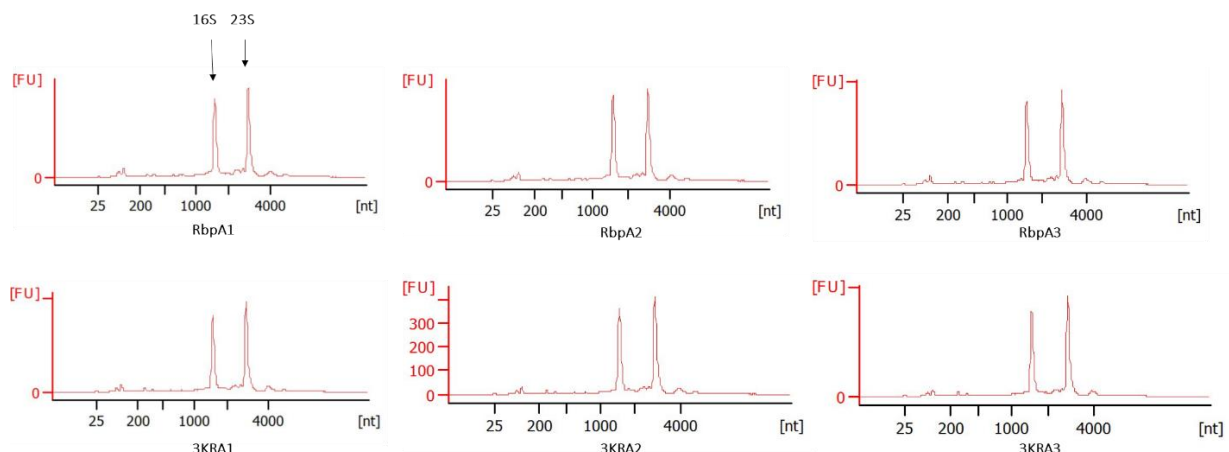


Figure 4.5. Bioanalyser results of the RbpA^{WT} and RbpA^{3KRA} RNA extractions from *M. tuberculosis*. Agilent 2100 Bioanalyzer NanoChip® systems, prokaryote total RNA nano assay. The 16S and 23S ribosomal subunits are shown for RbpA (M233 (pMV306::*rbpA*^{WT})) and 3KRA (M233 (pMV306::*rbpA*^{3KRA})) extracted from three independent biological replicates.

Strain	Volume ul	Concentration ng/ul	Mass ng	260/280	260/230
RbpA (1)	17	82.2	1233	2.11	2.23
RbpA (2)	17	65.6	984	2.10	2.36
RbpA (3)	17	67.2	1008	2.11	2.32
3KRA (1)	17	55.6	834	2.12	2.33
3KRA (2)	17	52	780	2.11	2.31
3KRA (3)	17	65	975	2.12	2.37

Table 4.1. Quality control results and quantification of the *M. tuberculosis* RNA samples. The RNA concentration for RbpA (M233 (pMV306::*rbpA*^{WT})) and 3KRA (M233 (pMV306::*rbpA*^{3KRA})) was measured using the Qubit RNA HS assay kit. The 260/280 and 260/230 ratios were measured using the Nanodrop spectrophotometer. This table represents the RNA samples sent for sequencing.

4.2.1 Sample preparation and library preparation at Oxford genomics centre

Library preparation of the 6 *M. tuberculosis* samples and sequencing was carried out at the Oxford Genomics Centre, The Wellcome Trust for Human Genetics (UK) using an Illumina HiSeq 4000. The samples were quality controlled again by the Oxford Genomics Centre before strand-specific paired-end sequencing was commenced. The gel image (Fig. 4.6) showed that all the samples had distinct bands for both ribosomal subunits and the RIN scores >9.6, demonstrating that the RNA quality is high. The electropherograms from the TapeStation did however indicate DNA contamination in all of the M233 (pMV306::*rbpA*^{3KRA}) biological replicates (Fig.4.6D) which

was not identified using the Bioanalyser. Therefore, all 6 of the samples were re-run through Qiagen RNeasy columns, repeating the on-column DNase treatment before sequencing.

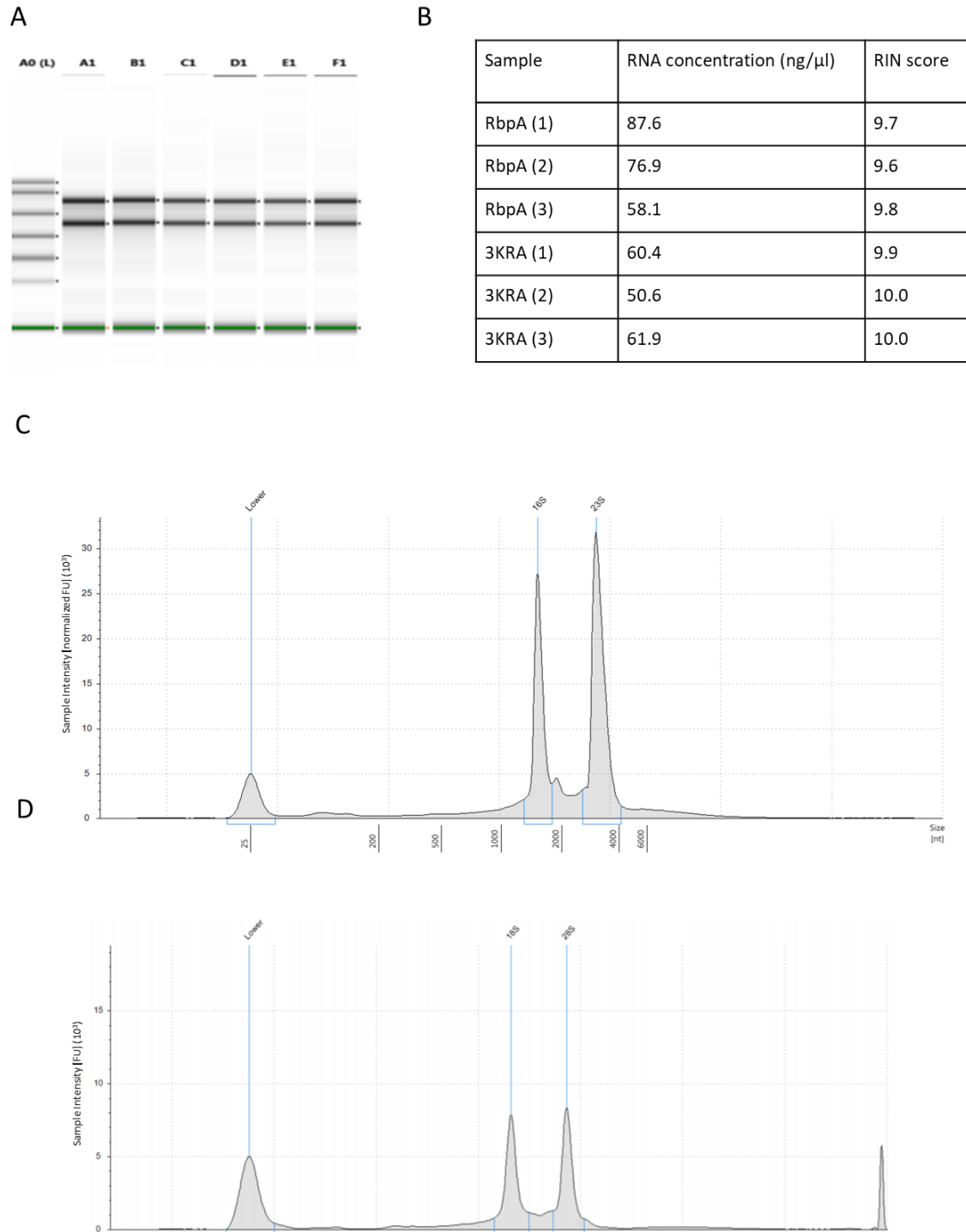


Figure 4.6. Agilent 4200 TapeStation quality control results for the *M. tuberculosis* BL RNA samples performed by the Oxford Genomics Centre. A. Reconstructed gel image of the RNA profiles. B. RNA concentrations and RNA integrity number equivalent (RINe) of the RNA samples. C. RNA profile of RbpA1 showing the ribosomal subunits 16S and 23S. D. RNA profile of 3KRA1 showing the ribosomal subunits 16S and 23S, the DNA contamination is shown upstream of the 23S subunit as it is higher molecular weight.

4.2.2 FastQC and trimming of the reads

The raw FASTQ files delivered from the Illumina HiSeq 4000 sequencer were processed as mentioned in described Results I: 3.1.7.2. The reads were trimmed from the 3' using Sickle (version 1.29) because sequencing at the periphery of the reads is usually poor, Table. 4.2 shows the number of paired end reads discarded. The quality of the RNA reads was improved after trimming (Fig. 4.7).

Sample	PE reads kept	PE reads discarded	% discarded
RbpA (1)	21637030	353	0.0016
RbpA (2)	25006892	607	0.0024
RbpA (3)	21412374	573	0.0027
3KRA (1)	24456552	657	0.0027
3KRA (2)	21656757	522	0.0024
3KRA (3)	24665222	791	0.0032

Table 4.2. The number of reads kept and discarded after trimming. Sequences were trimmed using Sickle (version 1.29). PE= paired end.

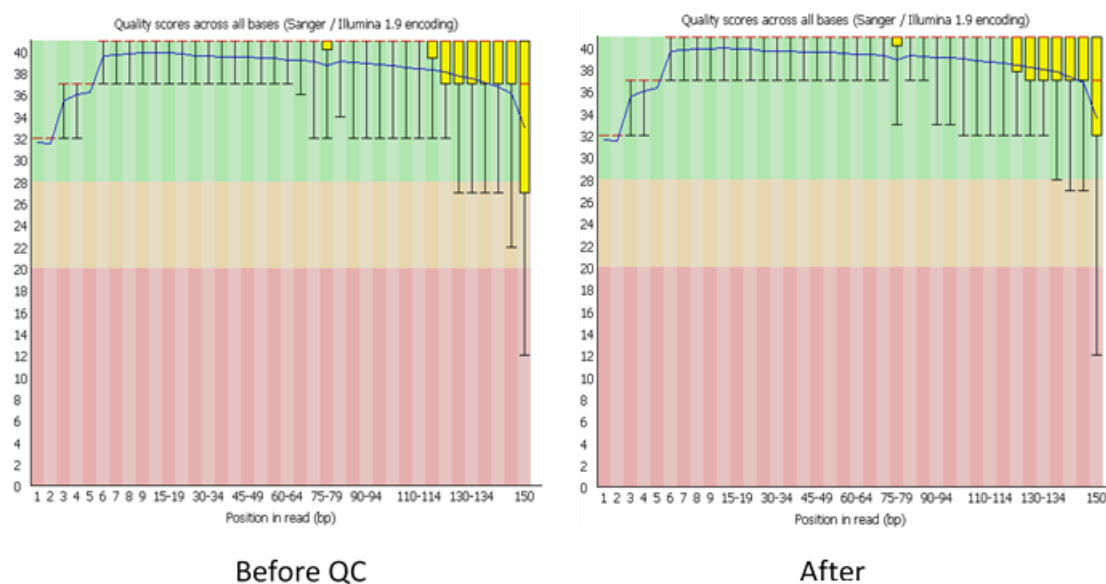


Figure 4.7. FastQC results of M233 (pMV306::rbpAWT) replicate 1 before and after trimming. FastQC (version 0.11.5) provides the base quality of the reads. (Left) before trimming, (right) after trimming using Sickel (version 1.29).

4.2.3 Mapping to the *M. tuberculosis* reference genome

The reference genome, *Mycobacterium tuberculosis* H37Rv, was downloaded from NCBI (NC_000962.3) in GTF and FASTQ formats, the GTF file contained information regarding the gene structure. The reads were aligned to the *M. tuberculosis* H37Rv reference genome FASTA file using HISAT2 (version 2.1.0). The reads for each sample demonstrated an overall alignment rate above 96% which eliminated any issues surrounding contamination (Table. 4.3). The number of reads mapping multiple times was also minimal (~0.15%). The discordant rate was low in all the samples which meant that both strands of the paired end sequencing mapped uniquely. The concordant rates of all the samples was ~92.6%, which suggested that the paired end reads map in close proximity to their pair.

Sample	Number of reads	Mapped reads (concordant)	Mapped reads multiple times	Overall alignment rate
RbpA (1)	21656757	19788345 (91.37%)	33617 (0.16%)	96.58%
RbpA (2)	21412374	19909299 (92.98%)	34004 (0.16%)	97.12%
RbpA (3)	24665222	22976737 (93.15%)	38986 (0.16%)	97.24%
3KRA (1)	21637030	20277461 (93.72%)	30141 (0.14%)	97.18%
3KRA (2)	24456552	22535682 (92.15%)	32815 (0.13%)	96.30%
3KRA (3)	25006892	23092678 (92.35%)	36141 (0.14%)	96.66%

Table 4.3. Overview of the sample reads using HISAT2. The table shows the total number of reads, mapped once, reads mapped multiple times and the overall alignment rate for 3 independent biological replicates of RbpA (M233 (pMV306::*rbpA^{WT}*)) and 3KRA (M233 (pMV306::*rbpA^{3KRA}*)). The average overall alignment rate is 94.75%. A very small number of the reads mapped to the genome multiple times, averaging 0.24%.

4.2.4 Assembling the transcriptome using Cufflinks

The mapped reads in SAM format were converted into their binary BAM files using SAMTOOLS (version 0.1.19), and the subsequent BAM files were sorted using the same index and aligned to the reference genome. Cufflinks (version 2.2.1; Fig. 4.9), a suite of tools was used to assemble (Cuffquant), merge (Cuffmerge), quantify and measure the differences (Cuffdiff) between M233 (pMV306::*rbpA^{WT}*) and M233 (pMV306::*rbpA^{3KRA}*; Results I; 3.1.7.4; Trapnell et al., 2011).

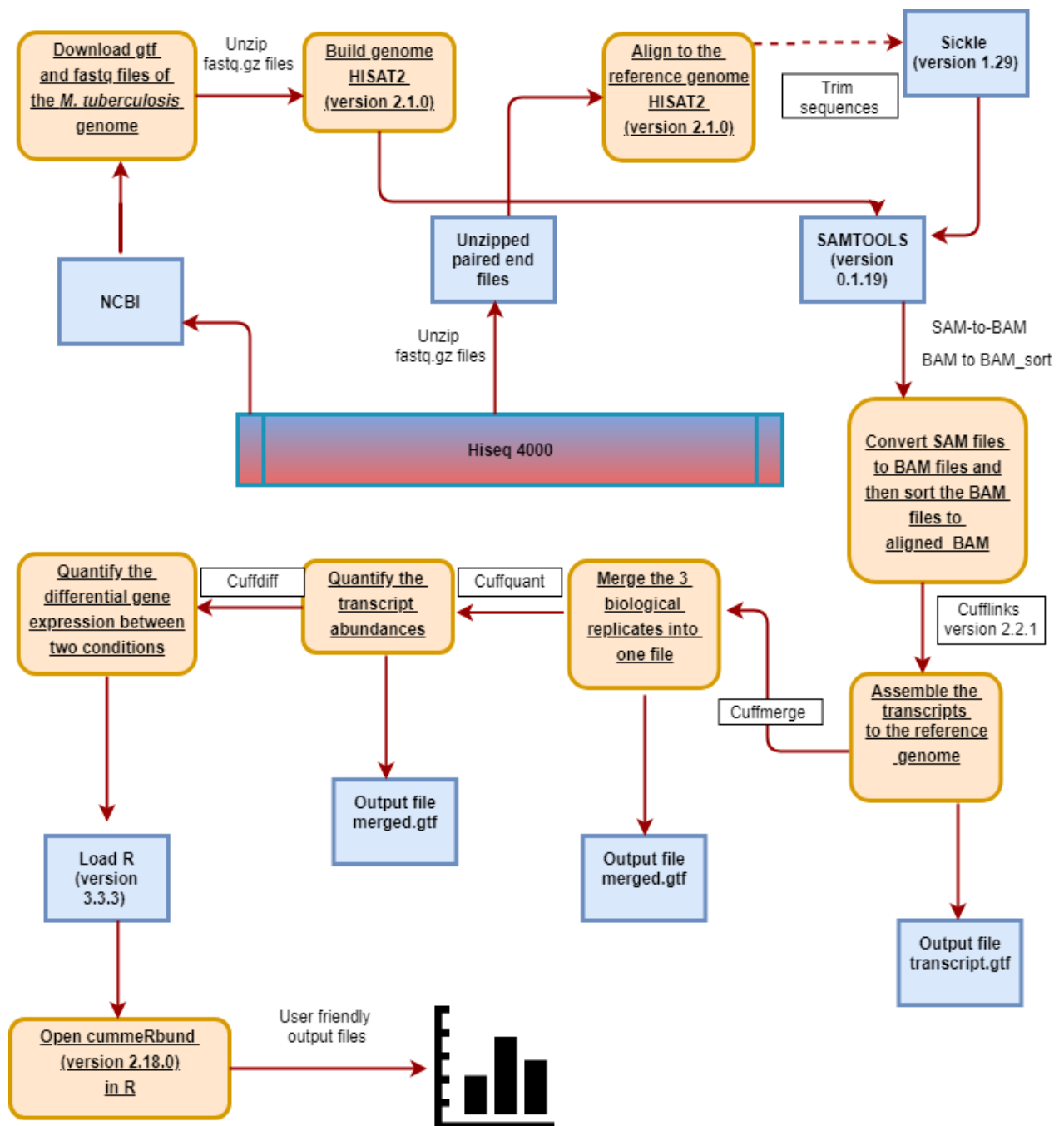


Figure. 4.8. Pipeline for the processing of the RNA-sequencing samples in *M. tuberculosis*. The arrows direct the pipeline steps. The white boxes represent the tools used to carry out the functions described in the orange boxes, whilst the blue boxes represent the input and output file formats.

4.2.5 Differential gene expression between M233 (pMV306::*rbpA*^{WT}) and M233 (pMV306::*rbpA*^{3KRA})

To identify the significantly differentially expressed genes, Cuffdiff from the Cufflinks application was implemented which provides one FPKM value for three biological replicates for each of the strains, M233 (pMV306::*rbpA*^{WT}) and M233 (pMV306::*rbpA*^{3KRA}; Fig. 4.9). To plot the individual replicate expression data, triplicate values were obtained by loading the differential gene expression list into R studio (version 1.0.153) using the CummeRbund package (version 2.0.0), this provided the individual FPKM values for each biological replicate. Cuffdiff was applied to identify significantly differentially expressed genes between M233 (pMV306::*rbpA*^{WT}) and M233 (pMV306::*rbpA*^{3KRA}). Cuffdiff presented genes that were statistically significant when the p value was lower than the FDR after Benjamini-Hochberg correction for multiple testing. 1236 genes (522 up-regulated and 714 down-regulated) were differentially expressed between M233 (pMV306::*rbpA*^{3KRA}) and M233 (pMV306::*rbpA*^{WT}) from a total of 4112 gene transcripts. The replicate values for all genes (FPKM) have been represented in a box plot and dendrogram (Fig. 4.9). The pattern of log₂ (FPKM) shown in the box plots was consistent across the replicates, whilst the dendrogram showed that the biological replicates cluster closely based on genetic background, therefore the samples were comparable. The significantly up-regulated and down-regulated genes in M233 (pMV306::*rbpA*^{3KRA}) compared to M233 (pMV306::*rbpA*^{WT}) were clustered and displayed as a heatmap, which highlighted good correlation of gene expression across biological replicates (Fig. 4.10).

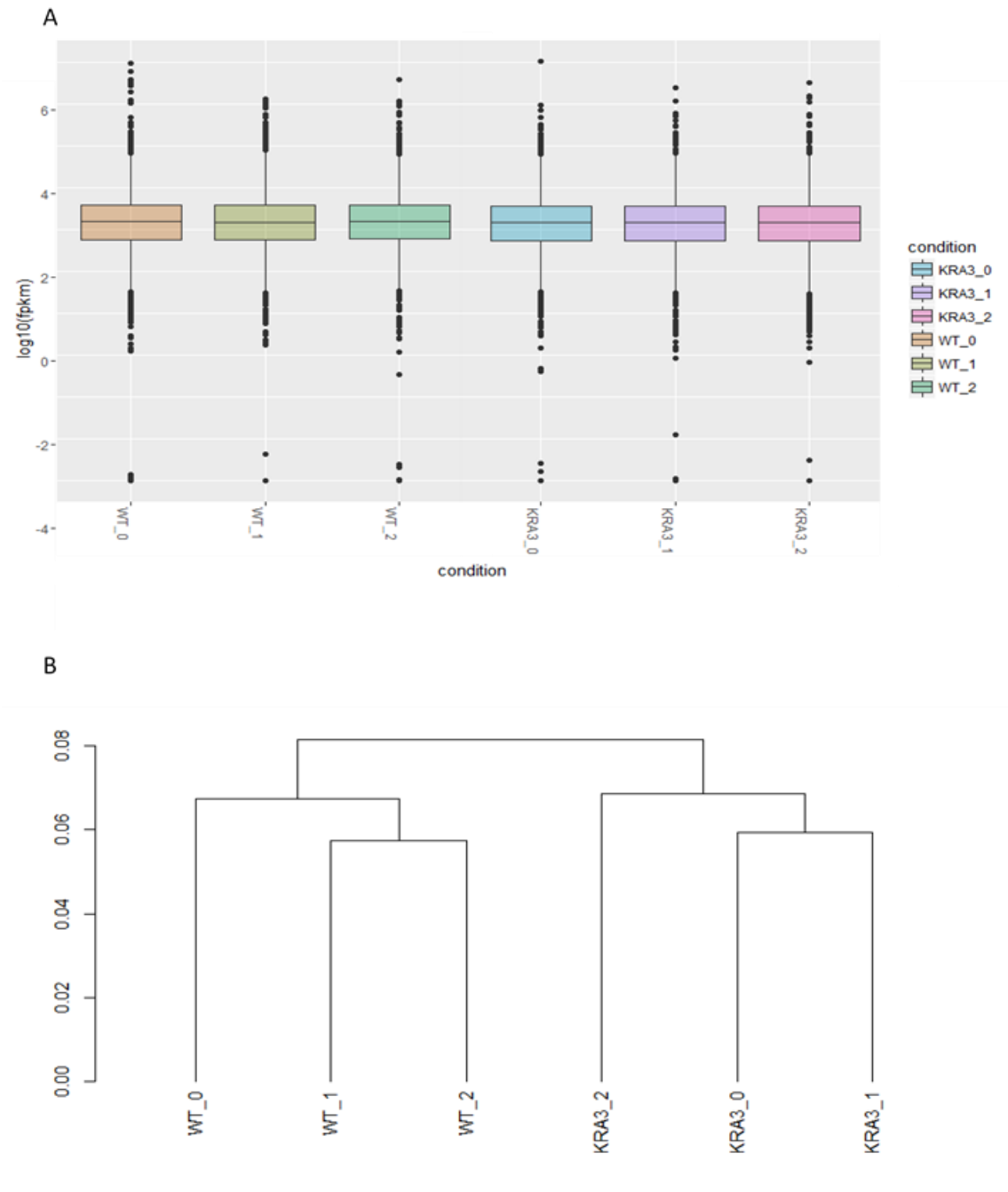


Figure 4.9. Comparison of the biological replicates FPKM for *M. tuberculosis*. (A) A boxplot showing the overall distribution of the gene expression of all three biological replicates for WT and 3KRA represented as log₁₀ FPKM for all 4112 genes. (B) A dendrogram showing the clustering of the biological replicates of, WT and 3KRA (KRA3) relative to each other based on gene expression levels (FPKM) of all 4112 genes.

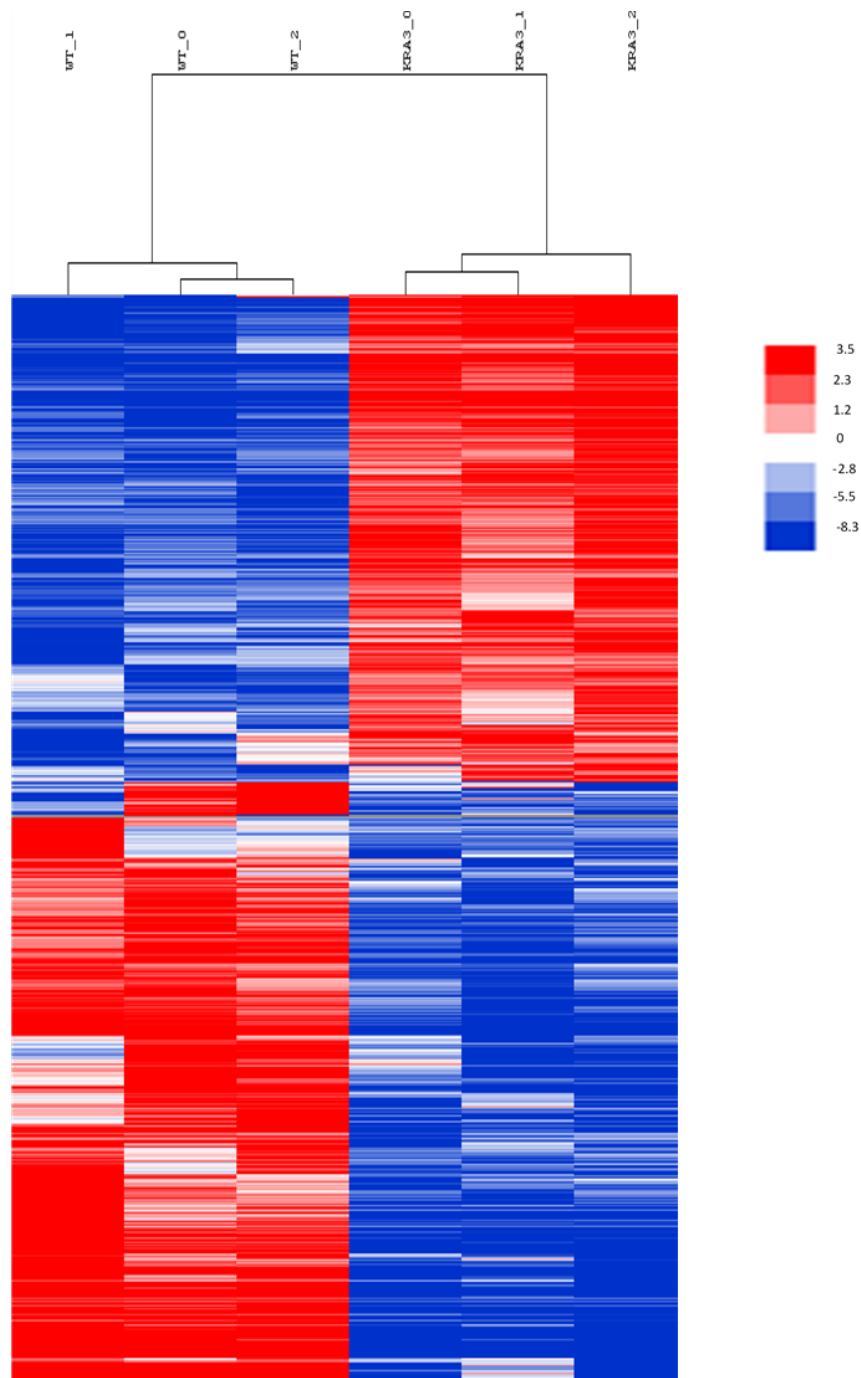


Figure 4.10. Differentially expressed genes in RbpA^{3KRA} compared to RbpA^{WT} in *M. tuberculosis*. Heatmap of significantly differentially expressed genes generated in Gene Cluster (version 3.0) using log transformed mean-centred data clustered by average linkage clustering and displayed with Java TreeView (version 1.1.6r4). Genes in red show log₂ fold change between 1.2-3.5 relative to the mean, genes shown in blue represent log₂ fold change from -2.8-8.3. The biological replicates cluster by genetic background as expected.

4.2.6 Minimal leaky expression from the repressed WT *rbpA* allele in M233 (pMV306::*rbpA*^{3KRA}) in the absence of PI

Empty vector, pMV306 that did not contain a plasmid- borne copy of *rbpA* did not grow on solid media, as expected, since *rbpA* is essential. However, this strain did grow in liquid culture raising concerns that the PIP repression system may be leaking expression in the absence of PI. To confirm this, the transcript levels of inducible *rbpA* versus the integrated pMV306::*rbpA*^{3KRA} copy in the M233 (pMV306::*rbpA*^{3KRA}) samples was assessed using Galaxy project tools. The FASTQ files for two biological replicates (was unable to load all three as a result of the size of the files) of M233 (pMV306::*rbpA*^{WT}) and M233 (pMV306::*rbpA*^{3KRA}) were uploaded into Galaxy (usegalaxy.org) and a 20 bp discriminatory sequence between the wild type *rbpA* and *rbpA*^{3KRA} was searched for using the Search in 'textfiles' (grep; Galaxy version 1.0.0). The output provided the number of reads which matched the two discriminatory 20 bp sequences. M233 (pMV306::*rbpA*^{WT}) averaged 804 reads for *rbpA*^{WT}, whilst the M233 (pMV306::*rbpA*^{3KRA}) averaged 71 reads for inducible *rbpA*^{WT} allele and 825 copies for integrated *rbpA*^{3KRA}, in the absence of the PI (Fig. 4.11). This showed that the level of *rbpA*^{3KRA} transcripts far exceeded the leaky expression of *rbpA*^{WT} transcripts in the absence of PI at a ratio of ~12:1. RbpA is an essential gene in *M. tuberculosis* therefore, it is possible that the low-level expression of *rbpA*^{WT} may be sufficient to maintain growth of the M233 (pMV306::*rbpA*^{3KRA}) comparable to M233 (pMV306::*rbpA*^{WT}) in liquid culture.

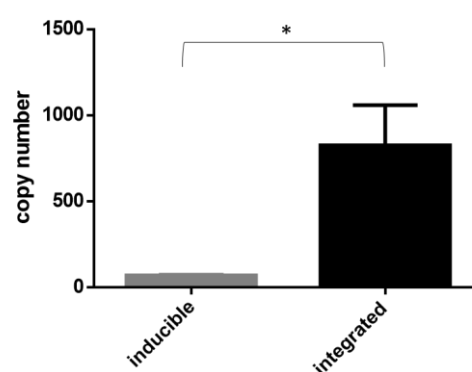


Figure 4.11. Quantified RbpA^{WT} and RbpA^{3KRA} reads from the M233 (pMV306::*rbpA*^{3KRA}) strain. Galaxy tool was used to search the reads for a discriminatory 20 base sequence for RbpA^{WT} or RbpA^{3KRA} to assay the stringency of the PI-inducible system. (Galaxy version 1.0.0). Results from

two biological replicates, error bars represent mean with standard deviation. Unpaired T-test p-value 0.0452.

4.2.7 Biological insights from transcriptome profiling of the RbpA BL mutant

Hypergeometric probability testing was performed to identify the biologically relevant patterns of gene expression by comparing the genes significantly induced and repressed in M233 (pMV306::*rbpA*^{3KRA}) vs M233 (pMV306::*rbpA*^{WT}) against 465 published and unpublished transcriptional signatures maintained in the Waddell lab. The list of genes as it stood were too large to obtain any real meaningful analysis from. The hypergeometric analysis (p value <0.05) was implemented to provide insight into the different pathways that RbpA may be implicated in.

4.2.7.1 Hypergeometric analysis of genes induced in M233 (pMV306::*rbpA*^{3KRA})

This section focuses on the signatures enriched in the 256 genes induced (1.5 fold change filter was applied to the list of 523 significantly induced genes) in M233 (pMV306::*rbpA*^{3KRA}) vs M233 (pMV306::*rbpA*^{WT}). A further filter for p value 0.005 was applied that resulted in a list of 16 significantly overlapping pathways, Table 4.4. Firstly, 190 genes from the RNA-seq data matched the non-essential genes from a transposon mutagenesis study which identified genes that are required for optimal *in vitro* growth in *M. tuberculosis* (Sasseti et al., 2003). RbpA has been shown to be present at almost all promoters therefore it is not surprising that a large number of genes transcribed by RbpA are non-essential. Secondly, 82, 54 and 27 genes overlapped with the Lsr2 binding sites (Gordon et al., 2010; Minch et al., 2015). Lsr2 is found in almost all sequenced Actinobacteria (Gordon et al., 2008) and has been known to bridge DNA together non-specifically (Chen et al., 2008). Therefore, it is expected that such a large number of genes are recognised as Lsr2 binding sites. Thirdly, a number of different datasets all investigating a non-culturable or persister state in *M. tuberculosis* were matched to the induced genes by M233 (pMV306::*rbpA*^{3KRA}). The first study showed that potassium depletion induced a non-culturable state which was resuscitated with the re-introduction of potassium (Salina et al., 2014). This dataset showed an overlap of 49 repressed genes for 2 days vs non-culturable (NC) state. Additionally, 18 genes were induced during -K vs log +K, these are

probably growth-related genes during the resuscitation of cells from a NC state. The second example shows that 40 induced genes in M233 (pMV306::*rbpA*^{3KRA}) map to genes induced in a nutrient starvation model at 4-96 h, during which growth is arrested (Betts et al., 2002).

Induced genes overlapping with Pathways	Number of overlaps	p-value
Non-essential genes. Sassetti et al., 2003	190	2.92E-08
Lsr2 binding sites (intensity ratio of 1.5) Gordon et al. 2010 PNAS	82	1.95E-07
Acid shock Induced <i>M.tuberculosis</i> , Golby et al 2007	19	1.94E-06
sigD putative regulon, Balazsi et al 2008	6	1.30E-05
All predicted and previously published secreted proteins Wiker et al., 2000	10	1.38E-05
Induced in log -K vs log +K t-test B+H p<0.05 >2 fold Salina 2014	18	1.69E-05
Repressed in Res 2d vs NC (high MPN:high CFU) Salina 2014	49	2.72E-05
PhoP regulon genes induced in WT v mutant Walters et al 2006	11	2.81E-05
Upregulated in vitro vs granuloma, pericavity, distant lung (Sup T14) Rachman et al., 2007	33	0.000109
Lsr2 binding sites (intensity ratio of 2.0) Gordon et al. 2010 PNAS	54	0.000185
Lsr2 repressed KO vs WT H37Rv in ANY Condition (Bartek et al., 2014)	27	0.000201
Predicted Secreted Proteins Top208-TM1 Gomez et al., 2000	10	0.000659
a BALBc at 14 days vs BALBc at 45 days Repressed Talaat et al 2007	3	0.00177
Nutrient Starvation Induced 4 or 24 or 96 h Betts et al., 2002	40	0.00228
sigL putative regulon, Balazsi et al 2008	3	0.00339
THP-1 Induced 24h, Fontan et al 2008	27	0.00374

Table. 4.4. Hypergeometric analysis of genes induced RNA-seq gene list comparing RbpA^{WT} and RbpA^{3KRA}. 256 significantly induced genes identified after applying a q-value cut-off 0.01, were run using the hypergeometric function. P-value cut-off (<0.05) was applied.

4.2.7.2 Hypergeometric analysis of genes repressed in M233 (pMV306::*rbpA*^{3KRA})

A 1.5 fold change filter was applied to the list of 714 significantly repressed genes resulting in a list of 373 genes for M233 (pMV306::*rbpA*^{3KRA}), were matched to the hypergeometric analysis and further filtered by applying a p-value 0.0001 cut-off. This filter resulted in 21 significantly enriched gene expression signatures (Table. 4.5). Firstly, a number of datasets again overlapped with genes involved in non-culturable *M. tuberculosis* states. Secondly, a study aimed to understand the transcriptional changes of host-microbe interaction showed an overlap of 128 genes induced by *M. tuberculosis* inside dendritic cells and macrophages (Tailieux et al., 2008). Thirdly, the DosR regulon identified by multiple studies was significantly enriched in the genes down-regulated by the RbpA BL mutations in *M. tuberculosis* (Section: 4.2.7.2.1).

Repressed genes overlapping with Pathways	Number of overlaps	p-value
NRP or Stationary phase Induced genes Voskuil et al., 2004	62	1.96E-37
Induced DC 1 or 4 or 18 h AND M 1 or 4 or 18 h Tallieux et al., 2008	128	2.60E-29
dosR Regulon Park 2003	36	1.53E-28
Induced NRP-2 vs log phase ttest p<0.05 B+H >1.5 fold Tudo et al., 2010	131	3.11E-25
Intracellular_Murine Macrophage Induced Schnappinger et al., 2003	104	4.95E-24
2.5 fold induced NRP2 v Aerobic ANOVA p0.05 B+H, Garton et al 2008	36	1.31E-17
dosR putative regulon, Balazsi et al 2008	22	4.10E-17
2.5 fold induced 70:30 v Aerobic ANOVA p0.05 B+H, Garton et al 2008	19	7.73E-17
RNI Induced genes Ohno et al., 2003	23	9.25E-17

NRP or Stationary phase Induced genes Voskuil et al., 2004	62	1.96E-37
THP-1 Induced 24h, Fontan et al 2008	27	0.00374
Induced DC 1 or 4 or 18 h AND M 1 or 4 or 18 h Tallieux et al., 2008	128	2.60E-29
Overexpressed Dendritic Cell v Macrophage Tallieux et al., 2008	66	1.10E-25
Induced NRP-2 vs log phase ttest p<0.05 B+H >1.5 fold Tundo et al., 2010	131	3.11E-25
dosR-dependent genes induced in vitro Kendall et al., 2004	26	3.57E-15
Low Oxygen Induced Bacon et al., 2004	31	1.46E-14
Bedaquiline Induced 30 mins 30x MIC top20 genes only, Koul et al 2013	15	1.76E-13
Repressed Res 192h vs NC ANOVA p<0.05 >1.5 fold Salina et al., 2014	132	2.36E-13
Induced by BTZ043 4h exposure at x10 MIC Makarov et al., 2009	26	2.66E-13

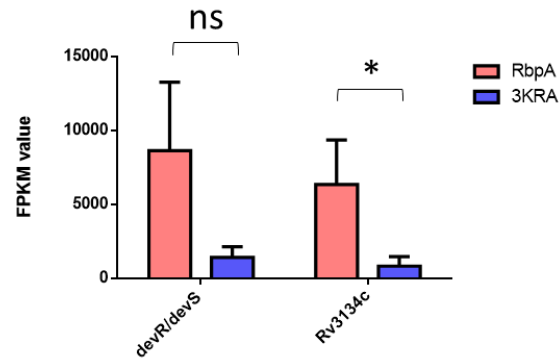
Table 4.5. Hypergeometric analysis of the genes repressed in $RbpA^{3KRA}$ compared to $RbpA^{WT}$. 373 genes after applying a q-value cut-off 0.01, were run using the hypergeometric function. p-value cut-off (<0.0001) was applied and some of the overlapping datasets are represented in this table.

4.2.7.2.1 Down-regulation of the DevR regulon by M233 (pMV306::*rbpA*^{3KRA})

The hypergeometric analysis revealed that the repressed genes in M233 (pMV306::*rbpA*^{3KRA}) overlapped with the DosR regulon. Therefore this section focuses on the status of the DosR regulon. The DevR (also known as DosR) regulon is induced by conditions that inhibit aerobic respiration and prevent bacillus replication, important for a non-culturable state (Sherman et al., 2001). DevR is a response regulator and DevS is a sensor kinase which binds to respiration-impairing gases such as NO and CO, and thus is activated upon oxygen deprived conditions (Kumar et al., 2007). There is a gene upstream of *devR*, Rv3134c which has no known function, the stop codon of Rv3134c provides a gap of 27 bp before the start codon of *devR*. The ORFs

of *devR* (Rv3133c) and *devS* (Rv3132c) overlap by one base, the TGA stop codon of *devR* shares the TG of the start codon of *devS*, ATG. It is believed that Rv3134c-*devR*-*devS* are co-transcribed as an operon (Dasgupta et al., 2000; Fig. 4.12B). The RNA-seq data of this study showed down-regulation of the *devR/devS* two component system in M233 (pMV306::*rbpA*^{3KRA}) by 6-fold (Fig 4.12A). The first gene of this co-transcribed operon, Rv3134c was also highly down-regulated by 7.7-fold.

A



B

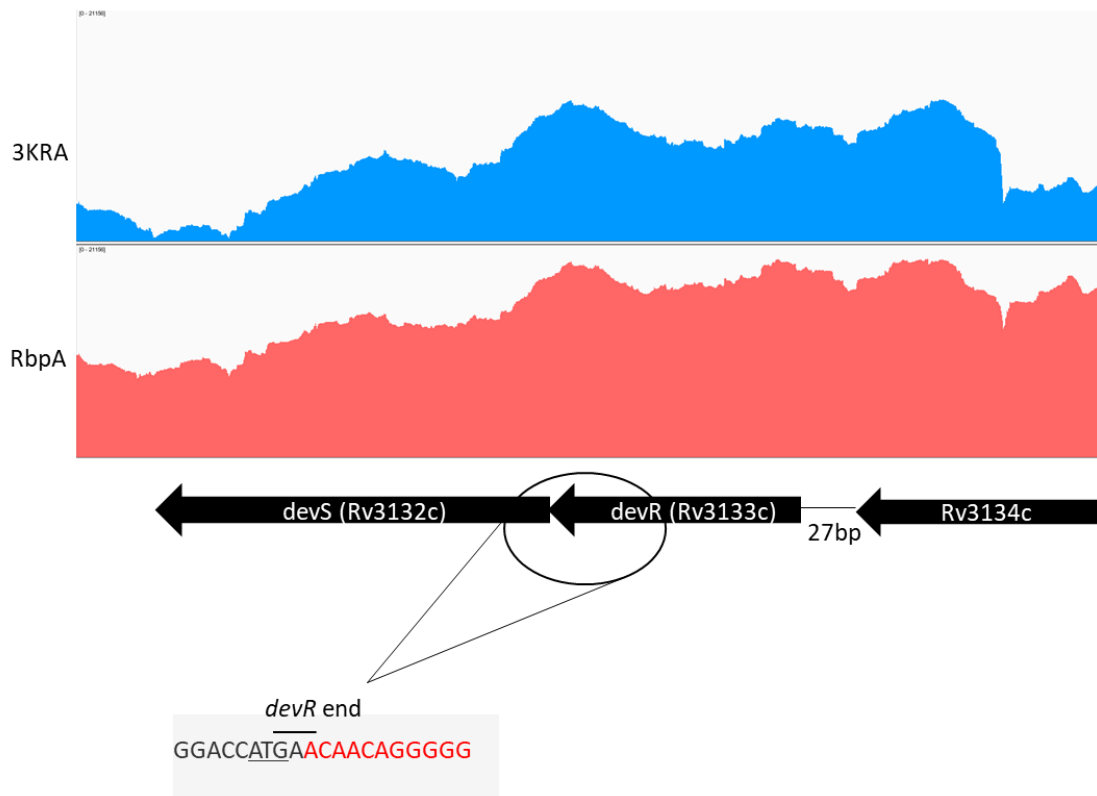


Figure 4.12. The *devS/devR/Rv3134c* operon visualisation. (A) displays the average FPKM values of the *devS/devR/Rv3134c* operon in 3KRA (blue) and WT (red); t-test comparing RbpA^{WT} and RbpA^{3KRA} Rv3134c p-value 0.0329 and *devS/devR* p-value 0.0554. (B) Visualisation of the FPKM values for three biological replicates for *devS/devR/Rv3134c* using IGV (version 2.4.5), the log scale has been normalised between the RbpA and 3KRA tracks, the overlap between *devS* and *devR* is a modification from Figure 5 from Dasgupta et al (2000).

Interestingly, a recent paper investigated the change in gene expression of various RbpA truncations, RbpA²⁸⁻¹¹⁴ (CD, BL and SID), RbpA⁷²⁻¹¹⁴ (BL and SID) and RbpA^{WT} in *M. smegmatis* (Hubin, et al., 2017a). These RNA-seq data was accessed through Gene Expression Omnibus database (accession number GSE89773) and processed through the RNA-sequencing pipeline (Figure 4.8). Three comparisons were made from this dataset, RbpA^{WT} vs RbpA²⁸⁻¹¹⁴, RbpA^{WT} vs RbpA⁷²⁻¹¹⁴ and RbpA²⁸⁻¹¹⁴ vs RbpA⁷²⁻¹¹⁴. The homologue of the *devR/devS* two-component system, MSMEG_3944/ MSMEG_3943 respectively is up-regulated in the RbpA⁷²⁻¹¹⁴ mutant by 3-fold, Fig 4.13. This suggested that RbpA BL may be required to facilitate in a response to hypoxia.

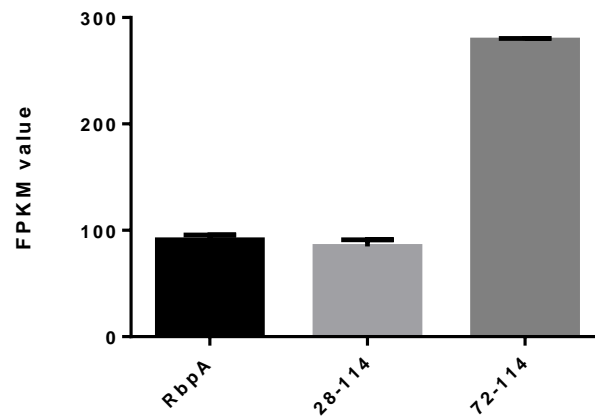


Figure 4.13. *devS/devR* FPKM values of the RbpA truncations, in *M. smegmatis*. FPKM values *devS/devR* (MSMEG_3944/ MSMEG_3943) in *M. smegmatis* with the Gene Expression Omnibus database (GEO: GSE89773; Hubin et al., 2017a) comparing RbpA^{WT}, RbpA²⁸⁻¹¹⁴ and RbpA⁷²⁻¹¹⁴. This data was compiled from two biological replicates.

Whilst the *devR/devS* operon itself was down-regulated in M233 (pMV306::*rbpA*^{3KRA}), it was important to identify the status of the DosR regulon. Studies have shown that in *M. tuberculosis* H37Rv exposed to low oxygen caused the induction of 47 genes from the DosR regulon in response to the hypoxic stress (Sherman et al., 2001). The RNA-seq data from this study confirmed that 36 out of these 47 genes of the DosR regulon were significantly down-regulated in M233 (pMV306::*rbpA*^{3KRA}) compared to M233 (pMV306::*rbpA*^{WT}; Fig. 4.14). The heat shock protein, *hspX* (Rv2031c) is a member of the DosR regulon and encodes α -crystallin (Acr), a major antigen found during latent infection of *M. tuberculosis* (Verbon et al., 1992). The

transcript levels of *acr* in the BL mutant were downregulated by 5-fold suggesting that the RbpA BL may be implicated in the host-microbe response during infection, particularly in the macrophage where bacilli encounter a multitude of stresses.

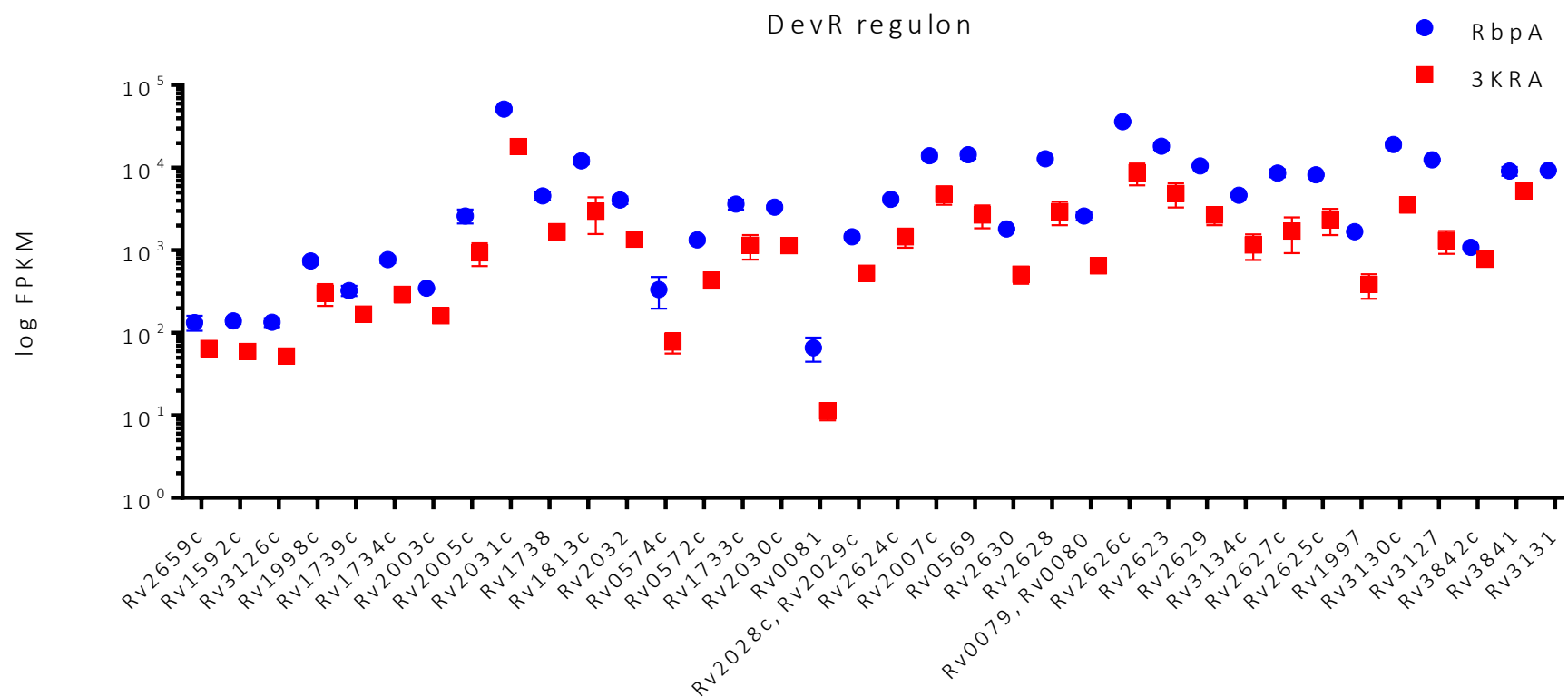


Figure 4.14. The DevR regulon genes down-regulated in M233 (pMV306::*rbpA*^{3KRA}) compared to M233 (pMV306::*rbpA*^{WT}). The FPKM values from two biological replicates of M233 (pMV306::*rbpA*^{WT}) and M233 (pMV306::*rbpA*^{3KRA}) have been plotted on a log scatter plot, the error bars represent mean with standard deviation.

4.2.7.2.2 RbpA^{3KRA} down-regulates the DosR target, class II ribonucleotide reductase

Class II RNRs modulate deoxynucleoside triphosphate biosynthesis in an oxygen independent manner and utilises adenosylcobalamin (vitamin B12) to generate radicals. Rv0570, *nrdZ* encodes for a class II RNR and transcription is dependent on *devR/devS* (Dawes et al., 2003). Deletion mutants of *nrdZ* have been shown to be dispensable for growth of *M. tuberculosis* (Dawes et al., 2003). The RNA-seq data showed that *nrdZ* was down-regulated in *rbpA*^{3KRA} by 8.8-fold (Fig. 4.15).

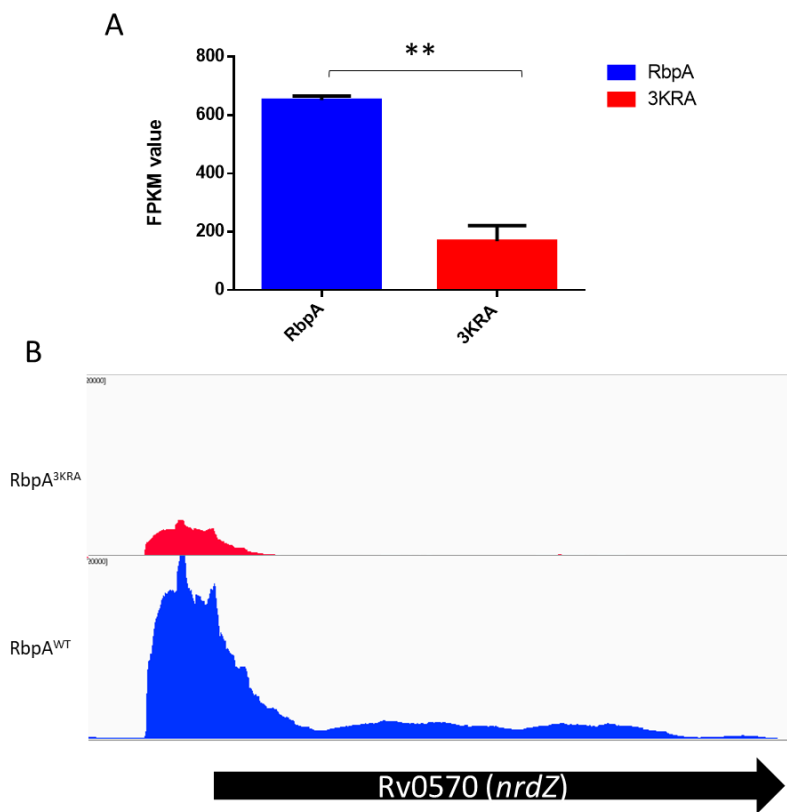


Figure 4.15. Down-regulation of *nrdZ* in by RbpA^{3KRA}. (A) Differential gene expression for *nrdZ* FPKM values for RbpA^{WT} and RbpA^{3KRA}, 652 and 167 respectively. The data is compiled from two biological replicates and the error bars represent mean with standard deviation (p =0.0065). (B) RNA-seq data visualised using IGV (version 2.4.10).

4.2.8 The RNAP subunits are up-regulated in M233 (pMV306::*rbpA*^{3KRA})

RbpA interacts extensively with σ^A and the RNAP β subunit (*rpoB*) during transcription initiation (Hu et al., 2012; Tabib-Salazar et al., 2013), so it was important to identify any changes in the expression levels of RNAP caused by M233 (pMV306::*rbpA*^{3KRA}). Thus far, this study has shown a down-regulation of the RNAP subunits for S401 (pRT802::*rbpA*^{3KRA}) in *S. coelicolor*. However, the differential gene expression comparing M233 (pMV306::*rbpA*^{3KRA}) to M233 (pMV306::*rbpA*^{WT}) revealed an up-regulation of *rpoB*, *rpoC* and *rpoZ* in *M. tuberculosis* by 2, 1.8 and 1.5-fold respectively (Fig. 4.16). The up-regulation of the RNAP subunits may suggest that RbpA plays some modulatory function for RNAP. Explanations for the up-regulation of *rpoB* specifically could be because the RbpA^{3KRA} is not fully functional and thus a general stress response is launched by the cell, or alternatively *rpoB* may be induced to increase the chances of holoenzyme formation that contains σ^A , because the expression levels of *sigA* remain the same in both strains.

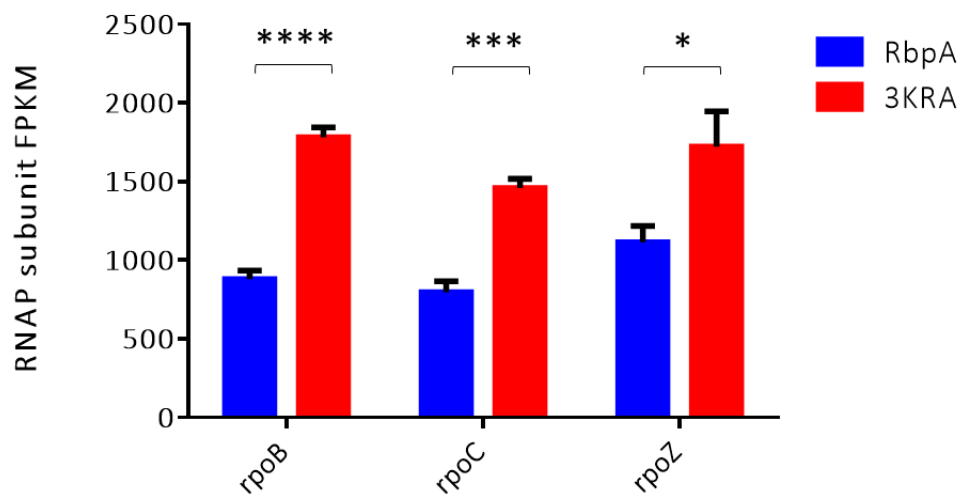


Figure 4.16. The up-regulation of RNAP subunits in M233 (pMV306::*rbpA*^{3KRA}). The data is compiled from three biological replicates and the error bars represent standard deviation with mean. The unpaired t-test p-value of *rpoB* <0.001, *rpoC* p value= 0.0002 and *rpoZ* p value= 0.0135 comparing 3KRA to WT.

The up-regulation of the RNAP subunits raised the question of why more RNAP is required in the cell for the M233 (pMV306::*rbpA*^{3KRA}) compared M233 (pMV306::*rbpA*^{WT}). Could this be a reflection of the greater rate of transcription/translation that is required during growth? Additionally, as bacterial cells increase in size with increasing growth rate the expression of ribosomal protein genes also increases to meet the demand for protein synthesis (Cortes & Cox, 2015). The RNA-seq results revealed that only five (out of 35) ribosomal protein genes were significantly differentially expressed between M233 (pMV306::*rbpA*^{3KRA}) and M233 (pMV306::*rbpA*^{WT}). Three genes, *rpsJ*, *rpmE* and *rpsO*, were up-regulated, and two genes, *rpsT* and *rpsP*, were down-regulated (Fig. 4.17). The *rpsJ* is one of the large classified ribosomal operons which consists of 11 ribosomal protein genes (Cortes & Cox, 2015). *rpsJ* was the only member of the operon that was significantly up-regulated, by 1.5-fold (Fig. 4.18), whilst the expression of *rplC*, *rplD*, *rplW*, *rplB*, *rpsS*, *rplV*, *rpsC*, *rplP* and *rpsQ* remained unchanged between RbpA^{3KRA} and RbpA^{WT}. The known TSS for the *rpsJ* operon was aligned to identify features of the promoter which would explain why *rpsJ* is the only member of the operon to be up-regulated (Cortes et al., 2013; Sachdeva et al., 2010). The alignment revealed that *rpsJ* was the only member of the operon to consist of a -10 and -35-element consensus sequence compared with the other members of the *rpsJ* operon (Fig 4.18B). Whilst the promoter alignment of the up-regulated ribosomal protein genes, *rpsJ*, *rpmE* and *rpsO*, and the down-regulated genes *rpsT* and *rpsP* revealed consistent conservation of the -10 and -35 promoter consensus sequence (Fig. 4.18C). From the promoter alignments, it was difficult to link RbpA and promoter dependency as the promoters from the up-regulated genes in M233 (pMV306::*rbpA*^{3KRA}) also consist of a well conserved -10 and -35 consensus sequence.

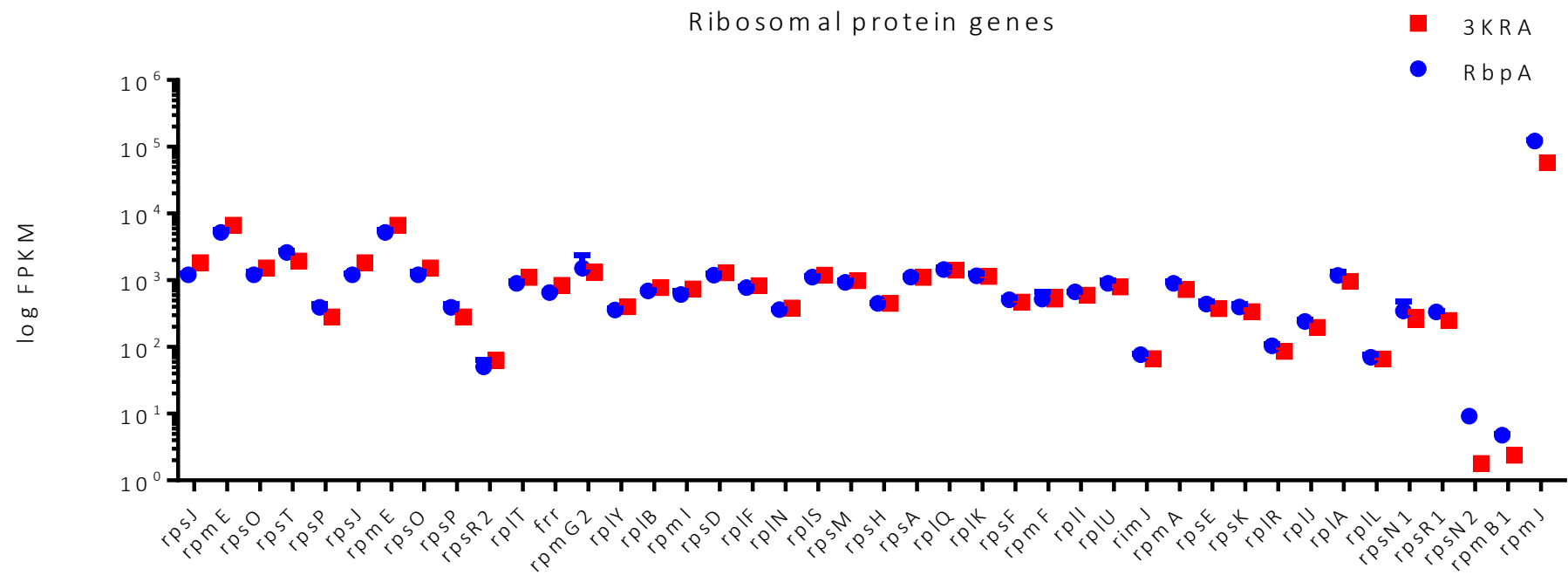
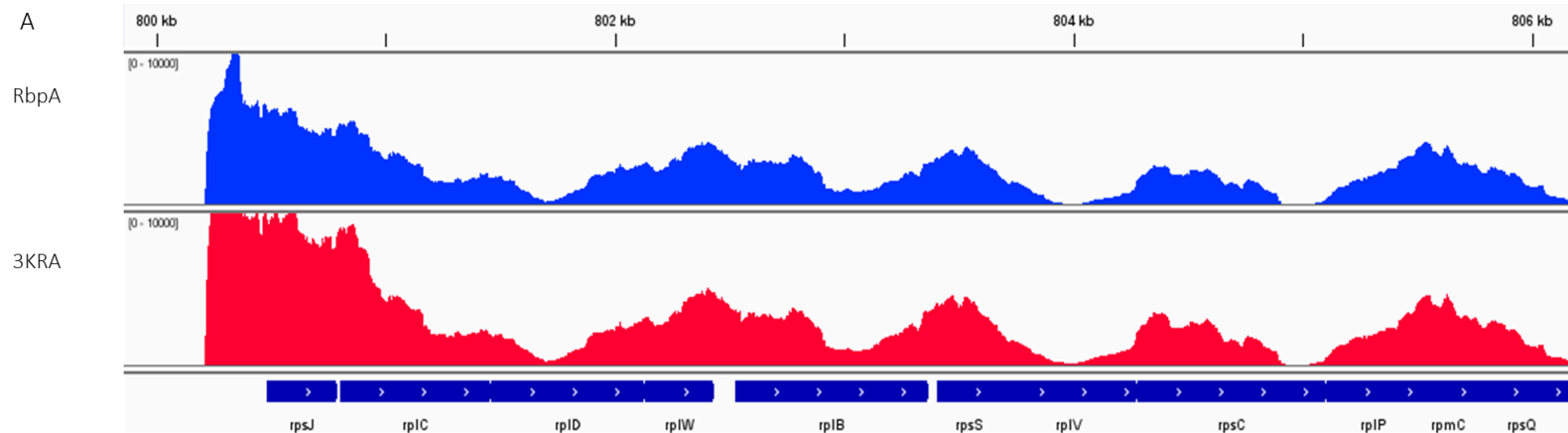


Figure. 4.17. The FPKM values for the ribosomal protein genes in *M. tuberculosis*. This data is compiled from three biological replicates and the error bars represent mean with standard error.



B

rpsJ CGAGCCAGCGGCGAT**TTGGCC**GAAGCGGTGATCTGGGGCA**TACTGT**TCAGG
 rplC TCGCGACAAGGTGTG**TGCGGG**CGACACGCCCAGCGCGGG**GCCGGT**GGAC
 rplD GCAATGACCGGGTGA**CCGTTC**TTAACCTTTTGGTGCATAA**GTCGAT**GCCG
 rplW CCTGACAGAACGTAA**ACAGGT**GCTGGTGGTCATCGGGCGC**AGCGAC**GAGGC
 rplB GACCGGAT---ACGG**CAAGCG**CAAGAGCACCAAGCGCGCC**ATCGTC**ACCCTGGC

TTGCGA N19 TANNNT
 -35 -10

C

rpsJ CGAGCCAGCGGCGAT**TTGGCC**GAAGCGGTGATCTGGGGCA**TACTGT**TCAGG
 rpmE ACGCATCTCGGTGT**TTGGGG**TGATAG-GTTGACCTGGCA**TAATCG**ATGCT
 rpsO CACCTGCGGCGT**TTGTGT**TTGGCGGTCAACCCGCTG**TAGACT**GT-CG
 rpsT CCACAGATTCCGCTT**TTGGGC**GAAAC-GTAACCGACTG**TAACCT**G-CAG
 rpsP CCCGCTGGGCGAA**TTACTG**CCCCAGCATCACGTTCTG**CACAAT**T-GGC

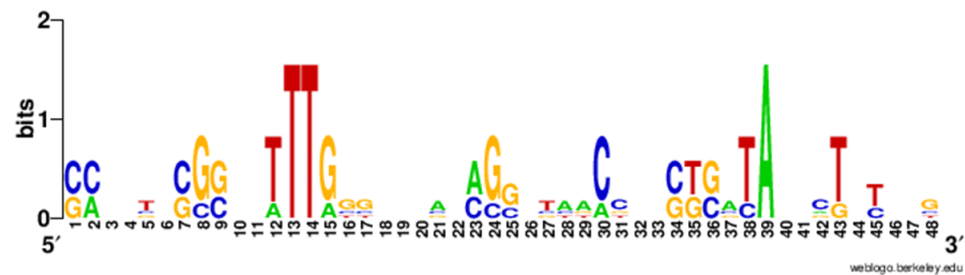


Figure 4.18. The *rpsJ* ribosomal protein operon. (A) The *rpsJ* ribosomal protein operon viewed in IGV (version 2.4.10), RbpA (RbpA^{WT}, blue) and 3KRA (RbpA^{3KRA}, red) showing the number of reads aligning at the *rpsJ* operon. (B) The TSS for the *rpsJ* operon genes (C) the promoters of the *rpsJ*, *rpmE* and *rpsO* which were up-regulated and *rpsT* and *rpsP* were down-regulated ribosomal protein genes using the Cortes et al., 2013, genome TSS annotations, the -10 and -35 elements are shown in bold and underlined, the extended -10 element is shown in bold. Motif conservation created using WEBLOGO (version 2.8.2), position 1 represents -48 and position 48 at the 3' presents +1 of the promoter. The *sigA*-dependant promoter consensus sequence is shown in bold (Sachdeva et al., 2010). +1 of the TSS is annotated in red. The promoter TSS were not available for *rpsS* and *rpIV*.

4.3 Identification of potential transcription factors involved in response to BL mutation in *M. tuberculosis*

A recent study in *M. tuberculosis* mapped transcription factor interactions by overexpressing transcriptional factors to define TF regulons using microarrays alongside ChIP-seq experiments to define TF binding patterns (Rustad et al., 2014). Hypergeometric probability was used to identify enriched TF regulons in the M233 (pMV306::*rbpA*^{3KRA}) vs M233 (pMV306::*rbpA*^{WT}) signature (log fold change=1 and p-value=0.01 filters were applied). Identification of large overlaps between these datasets could identify TFs that are specifically modulated by the RbpA BL.

4.3.1 Analysis of the overlapping genes induced in M233 (pMV306::*rbpA*^{3KRA}) to transcription factor over expression datasets

The 256 up-regulated genes in M233 (pMV306::*rbpA*^{3KRA}) were compared to TFOE database and sorted by significance (Ind.Sig p-value 0.05; Table 4.6). Several sigma factor regulons were significantly enriched in the genes induced after RbpA BL mutation, SigI, SigJ, SigG, SigF and SigH. SigI is one of ten ECF σ factors in *M. tuberculosis* and is up-regulated during stationary growth phase and during heat shock (Lee et al., 2012). SigJ is also expressed during stationary phase and $\Delta sigJ$ null mutants demonstrated increase susceptibility to

hydrogen peroxide (Yanmin et al., 2004). The RNA-seq data comparing M233 (pMV306::*rbpA*^{3KRA}) to M233 (pMV306::*rbpA*^{WT}) showed that *sigI* and *sigJ* were the only sigma factors up-regulated 4 and 2-fold respectively in M233 (pMV306::*rbpA*^{3KRA}). Although the number of transcripts aligning to the principal sigma factor, *sigA* was extremely high in both M233 (pMV306::*rbpA*^{WT}) and M233 (pMV306::*rbpA*^{3KRA}), FPKM values of 600 and 719, respectively (Fig. 4.19C). The SigE expression data from Results chapter I indicates that despite the high expression levels of HrdB in *S. coelicolor*, the levels of SigE which is an ECF sigma factor are unchanged, but more SigE associates with RNAP. Therefore it is quite possible that the ECF sigma factors up-regulated in the RNA-seq data and those that match the TFOE data may increase association with RNAP in *M. tuberculosis*.

TF	Name	TF.Ind	Ind.Sig
Rv3328c	sigI	3	2.64E-04
Rv3416	whiB3	255	1.56E-03
Rv0757	phoP	74	2.36E-03
Rv3736		5	2.39E-03
Rv0043c		11	3.85E-03
Rv1189	sigI	2	4.13E-03
Rv0182c	sigG	6	4.55E-03
Rv1358		27	6.29E-03
Rv0023		489	6.91E-03
Rv3286c	sigF	234	7.44E-03
Rv2034		28	7.58E-03
Rv3676	crp	43	1.84E-02
Rv3557c	kstR2	4	2.27E-02
Rv3260c	whiB2	104	3.35E-02
Rv3223c	sigH	95	3.99E-02
Rv3765c	tcrX	54	5.58E-02
Rv0212c	nadR	7	6.99E-02
Rv1994c	cmtR	16	7.92E-02
Rv3058c		136	9.59E-02

Table 4.6. Significantly enriched TF reulons in those genes up-regulated in M233 (pMV306::*rbpA*^{3KRA}) vs M233 (pMV306::*rbpA*^{WT}). TF.Ind refers to the number of genes induced by overexpression of the TF. Ind.Sig is the significance of the enrichment, p value 0.05.

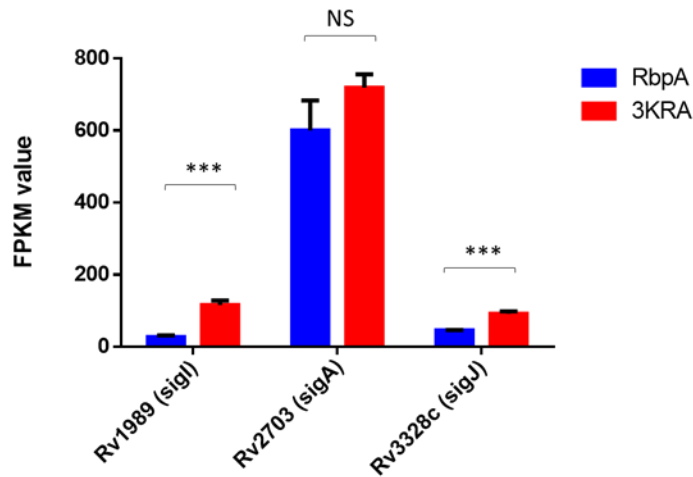


Figure 4.19. Significant up-regulation of SigI and SigJ in M233 (pMV306::*rbpA*^{3KRA}). Comparing *rbpA*^{WT} and *rbpA*^{3KRA} for *sigI* and *sigA* transcript levels using FPKM. Triplicate FPKM values for *sigI* (p value= 0.0003) and *sigA* (p-value = 0.0864) using GraphPad Prism (version 6.01).

4.3.2 The overlap of the down-regulated genes in M233 (pMV306::*rbpA*^{3KRA}) with the TFOE dataset

The 373 genes down-regulated in M233 (pMV306::*rbpA*^{3KRA}) were compared to the TFOE and sorted by significance (Rep.Sig p value 0.05; Table 4.7), resulting in a list of 11 genes. Again, DosR (DevR) appeared in this list which was been discussed with the down-regulated M233 (pMV306::*rbpA*^{3KRA}) hypergeometric analysis (Section; 4.2.7.2.1).

TF	Name	TF.Rep	Rep.Sig
Rv3249c		9	1.63E-03
Rv2884		134	2.05E-03
Rv0135c		14	1.00E-02
Rv1219c		39	1.09E-02
Rv3133c	dosR	43	1.84E-02
Rv1776c		11	2.95E-02
Rv0576		252	3.17E-02
Rv2788	sirR	72	3.96E-02
Rv2986c	hupB	6	5.21E-02
Rv3830c		1	6.43E-02
Rv2887		170	7.79E-02

Table 4.7. Significantly enriched TF regulons in those genes down-regulated M233 (pMV306::*rbpA*^{3KRA}) vs M233 (pMV306::*rbpA*^{WT}). TF.Ind is the number of genes repressed by overexpression of the TF. Ind.Sig is the significance of the enrichment, p value <0.05.

4.4 Discussion

4.4.1 RbpA^{3KRA} displays an attenuated growth phenotype on solid agar

This study focused on dissecting the function of the RbpA BL in *M. tuberculosis*. To study the effect of the RbpA BL mutations. RbpA BL mutants were constructed using a transposon inserted PI inducible system. The growth curve generated using OD readings in liquid culture over a period of 28 days, showed that the RbpA BL mutants were not attenuated. Checking the complementation of each of the RbpA BL mutant strains on solid media demonstrated, that M233 (pMV306::*rbpA*^{3KRA}) did influence growth on solid agar. This suggested that Pip may have been ineffective in completely repressing *pptr* in the absence of PI since the empty vector control containing no RbpA (an essential gene in *M.*

tuberculosis) still grew. RNA-seq was used to quantify the expression levels of *rbpA* expressed from the native PI inducible system in the absence of the inducer and the *rbpA*^{3KRA} allele in M233 (pMV306::*rbpA*^{3KRA}). This showed that there were 71 copies of *rbpA* vs 825 copies of *rbpA*^{3KRA} in this strain. As RbpA is essential in *M. tuberculosis*, even minimal expression of *rbpA* in the absence of the inducer could be enough to mask the true phenotype of the RbpA^{3KRA} allele. However, it is assumed that this same strain leaks *rbpA* expression in the absence of the inducer on solid media too, yet we observed an attenuated growth phenotype. To further investigate the non-culturability hypothesis, an MPN assay was performed alongside CFUs and OD readings. The OD readings consistently showed that the RbpA BL mutants did not influence growth rate in liquid media, whilst the CFU assays consistently demonstrated that M233 (pMV306::*rbpA*^{3KRA}) failed to grow as well as M233 (pMV306::*rbpA*^{WT}). Most strikingly, the MPN showed that the M233 (pMV306::*rbpA*^{3KRA}) required higher dilutions to identify the absence turbidity whilst the CFUs counts required lower dilutions for growth to be observed. This pattern of growth for high MPN vs low CFU has been associated with *M. tuberculosis* persister phenotypes (Mukamolova et al., 2010; Salina et al., 2014). However, the more likely explanation is that the Pip repression system is leaking expression of *rbpA* from the native loci, this is mainly because M233 (pMV306) does not grow on solid agar but does in liquid media. Most inducible systems in mycobacteria are known to be leaky and therefore it is difficult to estimate the extent of the leaky expression that masks the true phenotype of the RbpA BL mutants (Ehrt & Schnappinger, 2014). To address the possible instability of the transposon insertion, the RbpA-BL strains were sequenced after the initial transformations and after every growth curve. All sequencing confirmed the presence of the integration of RbpA BL mutants and the transposon insertion.

4.4.2 Impact of RbpA^{3KRA} mutation on *M. tuberculosis* transcriptome

Differential gene expression profiling comparing RbpA^{3KRA} to RbpA^{WT} demonstrated that the RbpA BL caused global transcriptional changes in the cell despite the absence of an attenuated growth phenotype in liquid culture (Hubin et al., 2017a). Hypergeometric analysis was performed to identify biological pathways enriched in the RbpA^{3KRA} vs RbpA^{WT}

gene list. This highlighted that RbpA was responsible for the induction of a large proportion of 'non-essential' genes (Sasseti et al., 2003). Both the up-regulated and down-regulated genes comparing the transcriptome of RbpA^{3KRA} vs RbpA^{WT} revealed a large overlap with persister/non-culturable phenotypes (Betts et al., 2002; Salina et al., 2014). This suggested that the RbpA^{3KRA} may cause global gene expression changes which signal for the bacterium to enter a non-replicating or persister phenotype. Additionally, it was found that the down-regulated genes in M233 (pMV306::*rbpA*^{3KRA}) overlapped with 43 DosR regulon binding sites. Closer inspection of the RNA-seq data from this study, highlighted that the co-transcribed *devR/devS/Rv3134c* operon and 36 DosR targets out of 47 were down-regulated in M233 (pMV306::*rbpA*^{3KRA}). Therefore, this pattern of transcriptional changes suggested that the RbpA BL residues could play a role in the *M. tuberculosis* response to hypoxic conditions, found in macrophages and lung lesions.

Additionally, the RNAP subunits were all induced by RbpA^{3KRA}, which could suggest that the loss of the full length RbpA signals a global stress response. The RNA-seq data also highlighted the up-regulation of multiple ECF sigma factors, thus more RNAP availability could result in the ECF sigma factors out-competing SigA for holoenzyme formation. Other classes of growth-related genes such as the ribosomal protein genes were investigated, this showed only 5 out of 35 to be affected by the BL residue mutations suggesting that this BL phenotype is more specific than a general slowing of growth. However, *rpsJ* was up-regulated in M233 (pMV306::*rbpA*^{3KRA}), upon closer inspection of the *rpsJ* operon, promoter alignments revealed that *rpsJ* was the only affected member of the operon that consisted of 'good' -10 and -35 promoter elements. Additionally a promoter alignment of all the ribosomal protein genes up/down-regulated in M233 (pMV306::*rbpA*^{3KRA}) consisted of highly conserved -10 (TANNNT) and -35 (TTGCGA) promoter elements. It may well be that RbpA controls promoters with specific -10 and -35 promoter features and could function as a repressor of some ribosomal protein genes. Other studies have demonstrated that differences between mycobacterial and *E. coli* promoter structures affects the overall stability of transcription initiation. Mycobacterial promoters require conservation of -10 and -35 promoter elements for the initiation of transcription, whilst *E. coli* promoter can activate transcription initiation from a conserved -10 promoter element (Agarwal & Tyagi, 2006). This study investigated the optimum -35 sequence required for optimal activity by base substitution of the -35 to -30 sequence, this showed that, TTGCGA, was required for

optimum activity (Agarwal & Tyagi, 2006). Additionally some studies have shown that a conserved -35 sequence is not required for transcription initiation in mycobacteria and streptomyces (Bashyam et al., 1996). Interestingly, *rpsJ* was the only member of the operon up-regulated by M233 (pMV306::*rbpA*^{3KRA}). Closer inspection of the promoter revealed that *rpsJ* was the only member of the operon to consist of a good -10 and -35 promoter consensus sequence. The RbpA^{3KRA} residues could increase promoter binding but may not necessarily increase transcription. Alternatively, the high binding strength of RbpA^{3KRA} may trap RbpA^{3KRA} onto the promoter and thus unable to activate transcription initiation. Similarly, RbpA^{3KRA} could activate a stress response such that the ribosomal protein genes with a highly conserved consensus -10 and -35 promoter sequence are induced.

4.4.3 Future work

One of the most striking results for this chapter has been the contradictory RbpA^{3KRA} phenotypic results in liquid and solid media, and discrepancies between CFU and MPN counts, one possibility may be that the RbpA^{3KRA} induces a non-culturable phenotype. If the RbpA BL mutants are entering a persister/ non-culturable phenotype, it will be important to show how these cells resuscitate back to a culturable state. A study investigated the effect of K⁺ deficiency on culturability and showed how a high MPN/low CFU count can be used as a measure of a non-culturable phenotype, found that *M. tuberculosis* deficient in K⁺ were capable of resuscitation when supplemented with K⁺ (Salina et al., 2014). The non-culturable transcriptome of K⁺ deficient cells show down-regulation of glycolysis and gluconeogenesis pathways, TCA cycle, fatty acid, mycolic acid biosynthetic pathways and up-regulation of genes involved in the catabolism of branched chain keto and amino acids. The RbpA BL mutant transcriptome affects multiple intermediate metabolism, respiratory and lipid metabolism genes. Alternatively, it is an assumption that organisms should behave the same in both liquid and solid, so the differences seen in this study may just be a result of the state of the culturing media. Finally, key indicator genes identified as significantly differentially expressed in RbpA^{3KRA} compared to RbpA^{WT} should be validated using an alternative technique such as quantitative RT-PCR.

Results chapter III: Investigating the role of RbpA in *S. coelicolor*

5.1.0 Overview

To date, one of the main established functions of RbpA has been contacting the -10 promoter DNA to help stabilise RPo complexes (Hu et al., 2014; Hubin et al., 2015, 2017). Data presented in Results chapter I suggests that the RbpA basic-linker-DNA interactions are not the essential role of RbpA, implying that RbpA plays additional functions. These potential roles might include: an essential contact point for other transcription factors; a chaperone for holoenzyme formation; or a stabilising role for σ^{HrdB} . This chapter focuses on the importance of the RbpA SID M85 residue, which was previously recognised as having a potentially important role and investigating the binding capacity of a σ^{HrdB} mutant that is unable to bind RbpA by ChIP-seq analysis. Finally, the global importance of RbpA on the transcriptome is investigated in a strain in which RbpA is severely depleted.

5.1.1 The importance of RbpA M85 residue interactions with σ^{HrdB}

Previously, the RbpA-SID was subjected to alanine site-directed mutagenesis (Tabib-Salazar et al., 2013), which revealed several key residues involved in interaction (e.g. R89/R90) with HrdB. Although not required for σ^{HrdB} binding, M85 was also shown to play an important role. *rbpA*^{M85A} only partially complemented the *S. coelicolor* S129 strain (Δ *rbpA*) and, unlike any other *rbpA* mutant, appeared to act dominantly by inhibiting the development of the wild-type strain (M. Paget, personal communication). Structural predictions suggested that the equivalent residue in *M. tuberculosis* (M84) makes non-polar interactions with σ^{A} -residue K334, which makes contact with the -12 promoter DNA at the beginning of the transcriptional bubble, and might itself make van der Waals interactions with the DNA backbone (Hubin et al., 2015). In support of a key role, the *M. tuberculosis* RbpA^{M84A} mutant was less active than WT in *in vitro* transcription assays (Hubin et al., 2015). Therefore, this section focuses on the amino acid substitution of M85 with neutrally charged alanine (A), leucine (L) and negatively charged glutamate (E), with the aim of potentially enhancing the dominant effect. These RbpA M85 mutant constructs were then introduced *into S. coelicolor* S401 to understand the phenotype in the absence of a possible interference from *rbpB*.

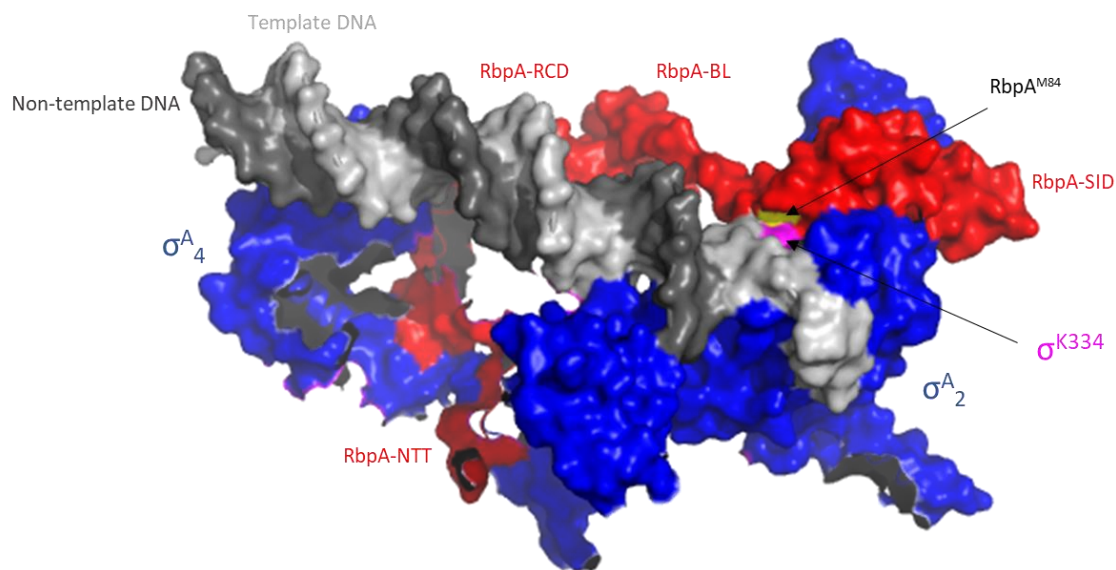


Figure 5.1. The cryo-EM structure of RbpA- σ^A complex in *M. tuberculosis*. Cryo-EM structure shows RbpA-SID complexed to σ^A_2 domain in *M. tuberculosis* and the RbpA^{M84} (RbpA^{M85} in *S. coelicolor*) is shown in yellow which interacts with σ^A - residue K334, shown in magenta. Accessed from Protein Data Bank using PDB ID: 6C04 and visualised and modified with PyMOL (version 1.3; Boyaci et al., 2018).

5.1.1.1 Construction of the RbpA C-terminal M85 mutants

Previous experiments on the RbpA-SID amino acid M85 mutants have been performed in *rbpA* null mutant backgrounds in which the paralogue *rbpB* has been deleted as it might mask the effects of certain alleles. Therefore, the M85 mutant alleles were introduced into S401 strain (A. Tabib-Salazar and M. Paget, unpublished), in which the only copy of *rbpA/rbpB* is an *rbpA* gene integrated in pIJ6902 such that it is induced by thiostrepton (Results I; 3.1.0). As mentioned previously (Results I: 3.1.0), the starting plasmid for constructing M85A, M85E and M85L alleles was pMT3000::*rbpA-NdeI*. This plasmid was used as a template to engineer the M85 mutant PCR products, *rbpA*^{M85A}, *rbpA*^{M85E} and *rbpA*^{M85L} using inverse PCR and confirmed by sequencing. The mutant alleles were amplified using primers that incorporated flanking HindIII sites 135 bp upstream of the *rbpA* start codon and replacing the stop codon, allowing them to be fused to the 3XFLAG epitope tag in pRT802::3XFLAG tag (Methods; Table 2.4), pRT802 integrates in the genome of *S.*

coelicolor, in this case at the ϕ BT1 attachment site (Gregory & Smith, 2003). Following confirmation of structure using SpeI and BamHI diagnostic digests, recombinant plasmids were transformed into *E. coli* ETR competent cells followed by conjugation into *S. coelicolor* S401 and S129. Thiostrepton was included in all sub-culturing media for the S401 strain to reduce the incidence of suppressor development.

5.1.1.2 Phenotype of the RbpA^{M85} mutants

The S401 strains were cultured with/without thiostrepton for five days at 30°C (Fig. 5.2A). As expected S401 (pRT802::*rbpA*^{WT}) grew equally well in the presence and absence of thiostrepton, whereas S401 (pRT802) formed very small colonies in the absence of the inducer. S401 (pRT802::*rbpA*^{M85L}) grew equivalently to the wild-type in the presence and absence of thiostrepton, whereas the M85A and M85E mutations each led to small colonies that overproduced ACT in the absence of thiostrepton, particularly the case for M85E. The expression of the alleles in *S. coelicolor* S129 background had only a minimal effect on growth, presumably as a result of *rbpB* suppressing the mutant phenotypes (Fig. 5.2B). The data revealed that replacement of M85 with a bulky hydrophobic residue did not significantly affect RbpA function, whereas a smaller uncharged residue (Ala) or a negatively charged residue (Glu) impacted on RbpA function leading to a growth defect and ACT overproduction. It is possible that a hydrophobic residue plays a key role in positioning K334 (*M. tuberculosis*) during promoter melting, although more investigation is required.

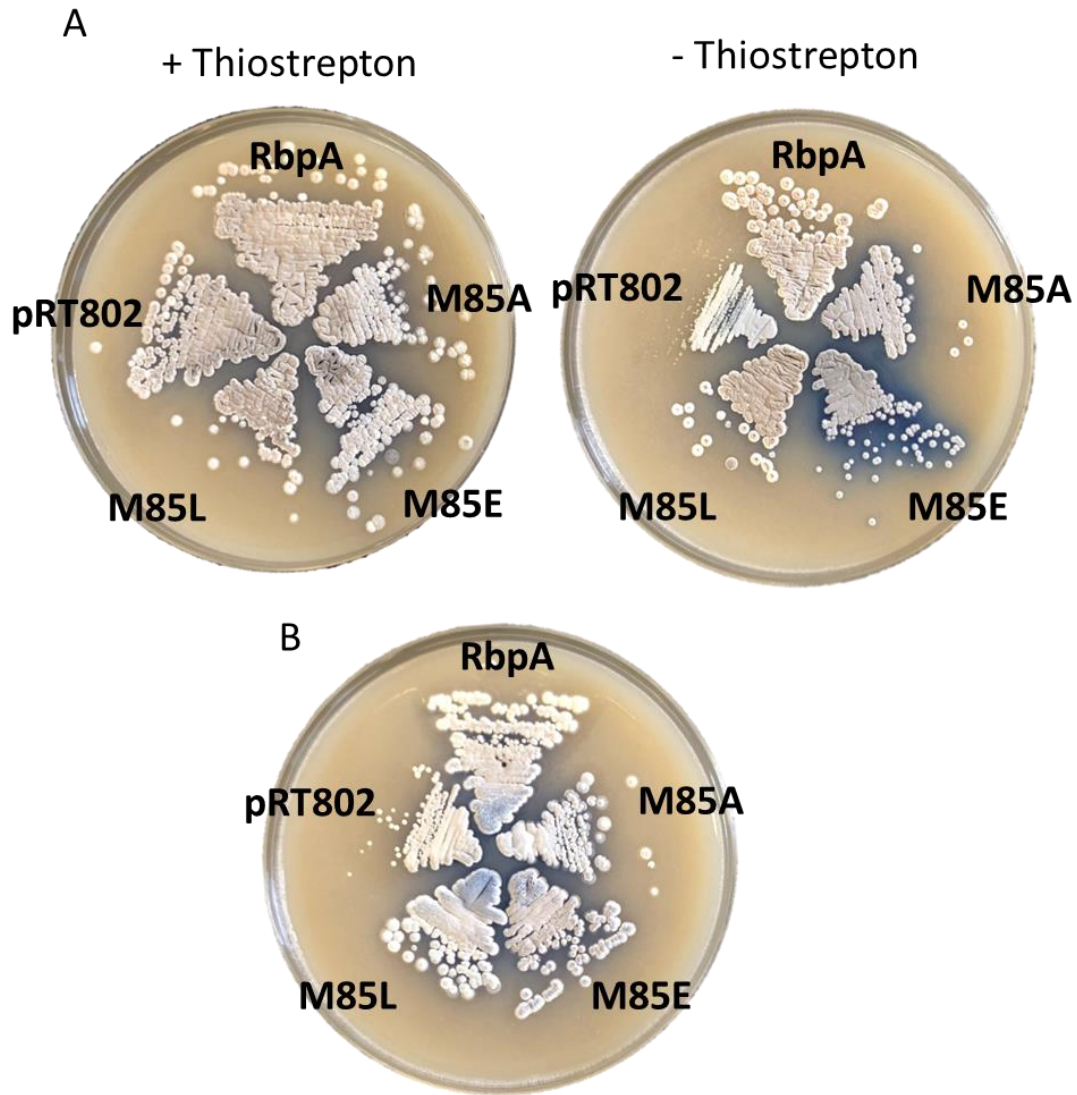


Figure 5.2. Phenotypes of the RbpA M85 mutants on MS agar. (A) S401 pRT802::*rbpA*^{WT} 3XFLAG, S401 pRT802 3XFLAG, S401 pRT802::*rbpA*^{M85A} 3XFLAG, S401 pRT802::*rbpA*^{M85E} 3XFLAG and S401 pRT802::*rbpA*^{M85L} 3XFLAG have been plated onto MS agar with (A, left) /without (A, right) 15 μ g/mL thiostrepton and incubated for five days. (B) S129 pRT802::*rbpA*^{WT} 3XFLAG, S129 pRT802 3XFLAG, S129 pRT802::*rbpA*^{M85A} 3XFLAG, S129 pRT802::*rbpA*^{M85E} 3XFLAG have been plated onto MS agar and incubated for five days.

5.2 Analysis of σ^{HrdB} mutants that cannot associate with RbpA

5.2.1 Overview

RbpA binds to group I sigma factors (*S. coelicolor* σ^{HrdB} and *M. tuberculosis* σ^{A}) and selected group II σ factors (*S. coelicolor* σ^{HrdA} and *M. tuberculosis* σ^{B}) but not others (*S. coelicolor* σ^{HrdC} and σ^{HrdD}). RbpA also does not interact with the more diverse group III (e.g. *M. tuberculosis* σ^{F}) and group IV proteins (e.g. *S. coelicolor* σ^{R}) (Hubin et al., 2015; Tabib-Salazar et al., 2013). A sequence alignment (Fig. 5.3) of the σ_2 domain revealed that σ^{HrdB} and σ^{HrdA} share 61% sequence similarity, whilst that between σ^{HrdB} vs σ^{HrdC} / σ^{HrdD} / σ^{R} is 50%, 40% and 27%, respectively. The sequence similarity between group I and some group II go some way in explaining why RbpA only binds to certain sigma factors.

The structure of *M. tuberculosis* σ^{A}_2 /RbpA revealed several amino acids in σ^{A}_2 that make important interactions with RbpA-SID (Hubin et al., 2015). These amino acids were substituted for the corresponding amino acids from *S. coelicolor* σ^{HrdC} , which as expected disrupted RbpA binding (Hubin et al., 2015). These findings provided an opportunity to investigate how σ^{HrdB} might behave *in vivo* when unable to interact with RbpA. Therefore, the equivalent substitutions were constructed in a 3XFLAG-tagged copy of σ^{HrdB} and the resulting effect on σ^{HrdB} binding across the *S. coelicolor* genome was analysed using ChIP-sequencing.

HrdB	EPGETENAGFVLSDEDEDDAP-AQQVAAA	GATADPVKDYLKQIGKVPLLNAEQEVELAKR	239
HrdA	HPG-----APVDVPEHPEPPP-PTRTESG	GPSSDLFRQYLRITGRIPLLSAAEVDLARR	124
HrdC	TARTRDDRRRA----	D YL QIG PLL AE EVRLATR	63
HrdD	-ARRKS-----	A DLVG YL RT	60
SigR	-----	-----VGPV-----TG	6
HrdB	TEAGLFAEDKLANSD----	KLAPKLRLEITIAEDGRRAKNHLEANLRLVVS	295
HrdA	VEAGLFAEELKRCSP----	GLDDRLALDRLVVLGRILAKRRLEANLRLVVS	180
HrdC	TEAGVRAREELETADTGEPAPT	RRRRRTLEETVHDGQAKDHMVRANLRLVVS	123
HrdD	TEAGVFARQVLEGYEET---	GADATREELQALIDESERAKDVFIRSNLRLVVA	117
SigR	TDAGTEHGQ--AEQPEGRGTGAESTAERSARFER	DALEFLDQMY-----SAALRMTR	56
HrdB	RGMLFLDLIQEGNLGLIRAVEKFDYTKGYKFSTYAT	WWIRQAITRAMADQARTIRIPVHM	355
HrdA	RGLTMLDLVOEGNLGLIRAVEKFDYARGYKFSTYAT	WWIRQAMSRALADQARTIRVPVHV	240
HrdC	RGLPLLDLVOEGNLGLIRAVEKFDHTKGEKFSTYAT	WWIRQAITRGLATHARTVRLPVHV	183
HrdD	SGLPLLDLIQEGNAGLIRAVEKFDYRKGEKFSTYAT	WWIRQAITRSIADQSRTIRLPVHL	177
SigR	NPADAEDLVQETYAKAYASFHQ-----	FREGTNLKAWLYRIINTFIN-----	99
HrdB	VEVINKLARVQRQMLQDLGREPTPEELAKELDMTPEKV	IEVQKYGREPISLHTPLGE-DG	414
HrdA	VELINRVVRVQRRMLQERGCEPTQEVAAHLDLAPER	VGEVLRLAQEPVSLHAPVGE-ED	299
HrdC	VEQLQKLAKVERKLRAGLDREPTTEEVAASGIDVKV	VWLRVGRDAVSLDTPVDE-TG	242
HrdD	VEELGRIRRVQREFNREHGREPEPAETAAELGSTPE	RVTDVLDWARDPVSLNMSVDD-EG	236
SigR	----SYRKKQREPQRSAAE-----	EIEDWQLARAESHMSTGLRSA	135

Figure 5.3. Amino acid sequence alignment of *S. coelicolor* Group I, II and the IV sigma factors.

The sequence alignment of conserved σ regions 1.2-2.3 in σ_2 domain. Sequence alignment performed using CLUSTALW multiple sequence alignment tool (version 1.2.2), identical amino acids have been shaded in black. The red arrows indicate the four amino acids from σ^{HrdB} that have been substituted with the corresponding σ^{HrdC} amino acids.

5.2.2.1 Construction of the σ^{HrdB} mutant for ChIP-analysis

Four amino acids not conserved (Fig. 5.3) between *S. coelicolor* σ^{HrdB} and σ^{HrdC} were identified for site directed substitutions:

- σ^{HrdB} E235/ σ^{HrdC} R59
- σ^{HrdB} K238/ σ^{HrdC} T62
- σ^{HrdB} L244/ σ^{HrdC} V68
- σ^{HrdB} F245/ σ^{HrdC} R69

The *hrdB* gene with 500 bp upstream DNA and a C-terminal 3XFLAG tag were inserted into the blunt ended EcoRV site in pBluescript II SK+ (created by Aline Tabib-Salazar) and was used as a template in this study. The σ^{HrdB} changes in conserved region 1.2, E235R and K238T, and the NCR, L244V and F245R, were generated using inverse PCR site-directed

mutagenesis using primers, HrdB-double_F and HrdB-double_R (Table 2.9); the resulting mutant was referred to as *hrdB^{4xR}*. The *hrdB^{WT}*3XFLAG and *hrdB^{4xR}* 3XFLAG genes were cloned into EcoRV-digested pSET152, and recombinants were digested with BamHI which confirmed the correct orientation of the inserts in pSET152, producing two fragments ~6320 bp and ~1342 bp. The confirmed pSET152::*hrdB^{WT}* and pSET152::*hrdB^{4xR}* constructs were transformed into *E. coli* ETZ and conjugated into *S. coelicolor* J1981 (*rpoC::6His*).

5.2.2.2 The HrdB mutant association with RNAP using western blot analysis

Initially, tests were performed to investigate whether the HrdB^{4xR} mutant could still associate with RNAP, and whether this was affected by a stress response involving the activation of an alternative σ factor. The best characterised σ based stress response in *Streptomyces* is the σ^R response to thiol stress (Kallifidas et al., 2010) which involves the induction of >100 genes in response to thiol oxidation by e.g. diamide. Interestingly, *rbpA* is a member of the σ^R regulon and is therefore also induced in response to diamide stress (Newell et al., 2006). Although it is not known why *rbpA* is controlled by σ^R , one possibility is that an increased expression of *rbpA* might have an important role in maintaining σ^{HrdB} RNAP levels in the face of increased competition. Therefore, the association of the HrdB^{4xR} mutant was investigated under conditions where σ^R levels were induced, providing increased competition with σ^{HrdB} . In this assay the J1981 *rpoC::6His* allele enabled the rapid purification of RNAP via immobilised metal affinity chromatography, followed by the immunodetection of associated FLAG-tagged HrdB^{WT} or HrdB^{4xR}. Firstly, spores for J1981 (pSET152::*hrdB^{WT}* 3XFLAG) and J1981 (pSET152::*hrdB^{4xR}* 3XFLAG) were cultured to OD₄₅₀ 0.8 (Methods; 2.4), 1 mM diamide was applied to the cultures and 10 mL samples at time points 0, 20, 40 and 60 min were harvested. The his-tagged RNA polymerase in these strains were purified using a cobalt mini-column, and the eluates tested by western blotting using an anti-FLAG monoclonal antibody (Methods; 2.8.3) to detect FLAG-tagged σ^{HrdB} . The western blot was normalised by dividing each HrdB signal by the RNAP β subunit, then divided by HrdB^{WT} 0 min. Firstly, the time point 0 min for HrdB^{WT} and HrdB^{4xR} showed different amounts of immunodetection, this casted doubt on the validity of this western

analysis. Upon diamide induction, it was expected that the competition between σ^{HrdB} and σ^{R} would increase resulting in less HrdB associating with RNAP in the presence of oxidative stress. However, the amount of HrdB detected in HrdB^{WT} samples was generally the same, which could be because RbpA is induced by the SigR regulon (Newell et al., 2006).

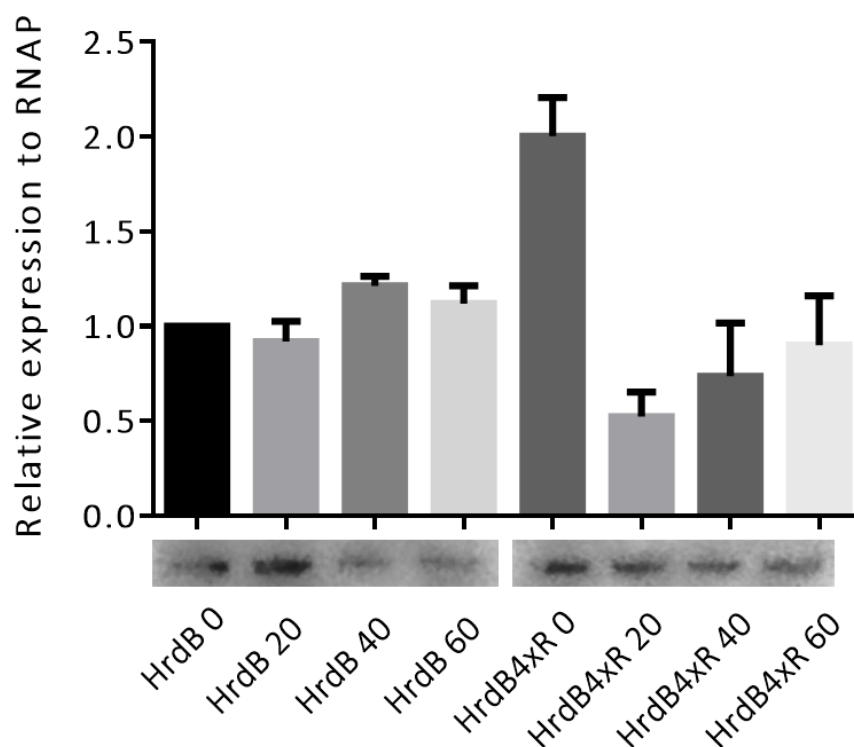


Figure 5.4. Western blot analysis of HrdB association with RNAP upon diamide stress. Spore stocks were normalised and cultured in YEME supplemented with glucose to OD₄₅₀ 0.8, 1 mM diamide was added to the cultures and at time points 0, 20, 40 and 60 min, 10 mL samples were collected. The cell lysate was transferred to the cobalt column following manufacturer's instructions. The primary antibody was anti-mouse anti-FLAG whilst the secondary was anti-mouse anti-HRP. The data has been normalised firstly to the RNAP subunit control and then to HrdB time point 0 min. This figure is the result of two biological replicates and imaged using LICOR.

5.2.2.3 Chromatin immunoprecipitation of HrdB^{WT} and HrdB^{4xR}

Chromatin immunoprecipitation (ChIP) is a technique used to study protein-DNA interactions. This involves crosslinking all protein-DNA complexes followed by precipitating the protein of interest with an appropriate antibody and then analysing the DNA fragments purified, typically through qPCR or NGS. This section focuses on the ChIP-sequencing of DNA fragments immunoprecipitated by the anti-FLAG antibody designed to isolate HrdB FLAG tagged. The strains J1981 (pSET152::*hrdB*^{WT} 3XFLAG) and J1981 (pSET152::*hrdB*^{4xR} 3XFLAG) were cultured in 80 mL YEME to mid exponential phase (OD₄₅₀ = 0.8-1). The ChIP procedure is detailed in Methods Section 2.10.

5.2.2.3.1 RbpA is not required for σ^{HrdB} interaction with SCO5281 promoter

HrdB^{WT} and HrdB^{4xR} ChIP samples were initially tested for the σ association with promoter DNA using qPCR to confirm that the C-terminal FLAG tag did not prevent promoter association and provide initial assessment of the effect of preventing RbpA binding. A large gene, SCO5281 (3819 bp) was chosen to allow resolution between the promoter and reading frame probes considering that the ChIP experiment involves shearing DNA to 300-500 bp fragments. SCO5281 is annotated as a putative 2-oxoglutarate dehydrogenase, an important TCA cycle enzyme which converts 2-oxoglutarate to succinyl semi-aldehyde. The 'promoter' forward primer (3SCO5281_F) was designed 21 bp upstream of the TSS (Jeong et al, 2016) and the reverse primer (3SCO5281_R) was designed 73 bp downstream of the TSS (Table 2.9). The 'mid-frame' forward primer (8SCO5281_F) was designed 2475 bp downstream from the TSS and the reverse primer (8SCO5281_R) was designed 1324 bp upstream of the stop codon (Table. 2.10). qPCR revealed a 6.9- and 4.6-fold enrichment at the promoter for HrdB^{WT} and HrdB^{4xR}, respectively (Fig. 5.5). The copy number was only slightly reduced in the HrdB^{4xR} compared to HrdB^{WT}, providing further evidence that the HrdB^{4xR} can associate with RNAP holoenzyme in the absence of RbpA and can bind to promoter DNA.

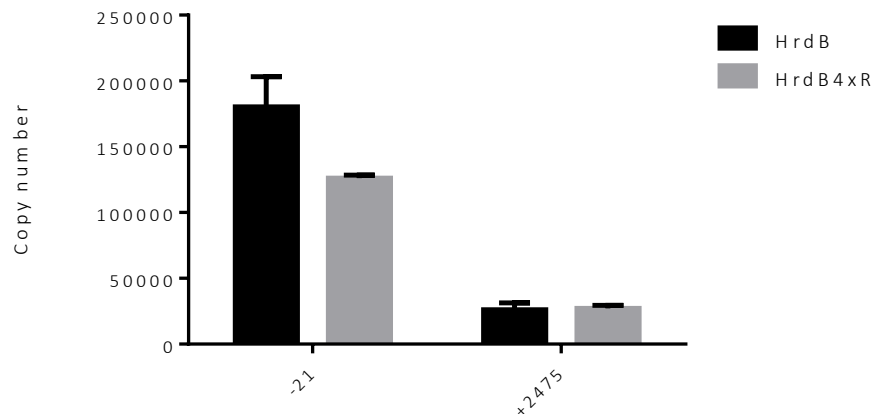


Figure 5.5. RbpA is not required for σ^{HrdB} interaction with the SCO5281 promoter. -21 represents 21 bp upstream of the TSS of SCO5281 and +2475 represents 2475 bp downstream of the TSS. This data is compiled from 2 biological replicates. Error bars represent mean with standard deviation.

5.2.2.4 Preparation of the ChIP-seq library

Samples were quality controlled by quantifying the DNA concentrations using the Qubit DNA HS assay kit following manufacturer's instructions. Additionally, the quality of the DNA was measured using the NanoDrop spectrophotometer (Table 5.1). The immunoprecipitated and input control DNA was sent to Novogene (Hong Kong) for library preparation (Fig. 5.6) and sequencing. The DNA fragments were extracted to construct the Illumina sequencing library. The ends of DNA fragments were end repaired and ligated to an adaptor (Table 2.10). The DNA fragments were run on a gel and fragments between 100-500 bp and extracted from the gel and PCR amplified for 16 cycles. The amplified PCR products were quality tested and sequenced. 8 samples were combined and sequenced using the Illumina HiSeq 2500 platform with 50 bp single-end reads via multiplex sequencing, the de-multiplexed sequences were received as FASTQ files for each ChIP sample.

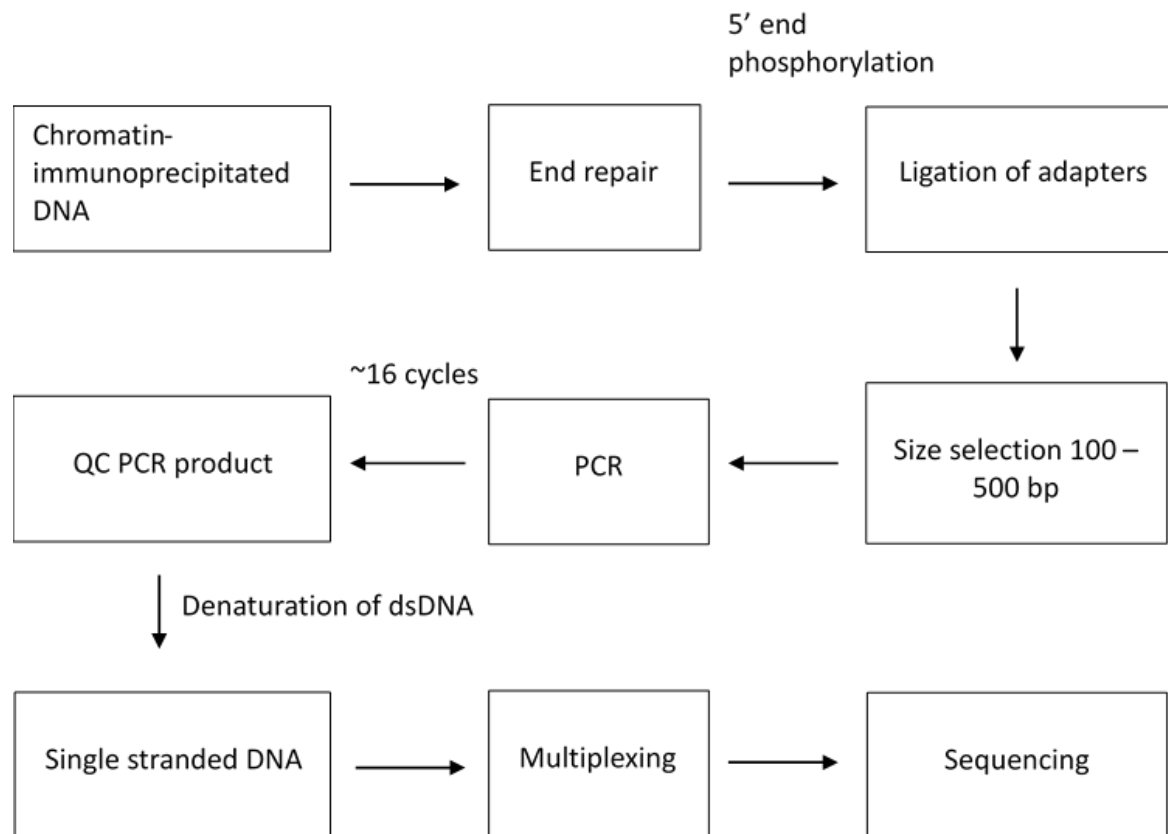


Figure 5.6. Pipeline for the library preparation of ChIP-seq samples. The samples were sequenced externally by Novogene (Hong Kong) using Illumina HiSeq SE50. The DNA fragments ends were repaired and ligated to an adaptor. The samples were run on a gel to enable size exclusion of any fragment above 100-500 bp. PCR of the fragments was enabled, and the PCR product was QC'd, denatured to ssDNA, multiplexed and sequenced. The sequenced files were received in FASTQ format.

Sample	ng/ μ l	Volume (μ l)	Total amount (ng)	260/280	260/230
WT1	0.634	16	10.144	1.1	0.68
WT2	0.134	16	2.114	1.29	0.63
WT3	0.27	16	4.32	1.23	0.57
HrdB4xR1	0.646	16	10.336	1.25	0.54
HrdB4xR2	0.506	16	8.096	1.51	0.64
HrdB4xR3	0.55	16	8.8	1.38	0.67
WT3 IPC	82	16	1312	1.82	1.53
HrdB4xR1 IPC	82.6	16	1321.6	1.84	1.87

Table 5.1. Quality control of the ChIP samples in-house. The ChIP samples were QC'd using FastQC Read Quality reports (Galaxy version 0.69). WT= *hrdB^{WT}*, mutant = *hrdB^{4xR}* and IPC= input control.

5.2.2.4.1 Bioinformatic analysis of ChIP-seq data

The FASTQ files were processed using the Galaxy tool (<https://usegalaxy.org>). The FASTQ files were first converted from Illumina to Sanger format using FASTQ Groomer (Galaxy Tool version, 1.0.4). The FASTQ files were quality control checked to ensure the sequencing was successful using FASTQC (Galaxy Tool version, 0.63) and highlights any problems with the read quality before any further analysis is undertaken (Fig 5.7).

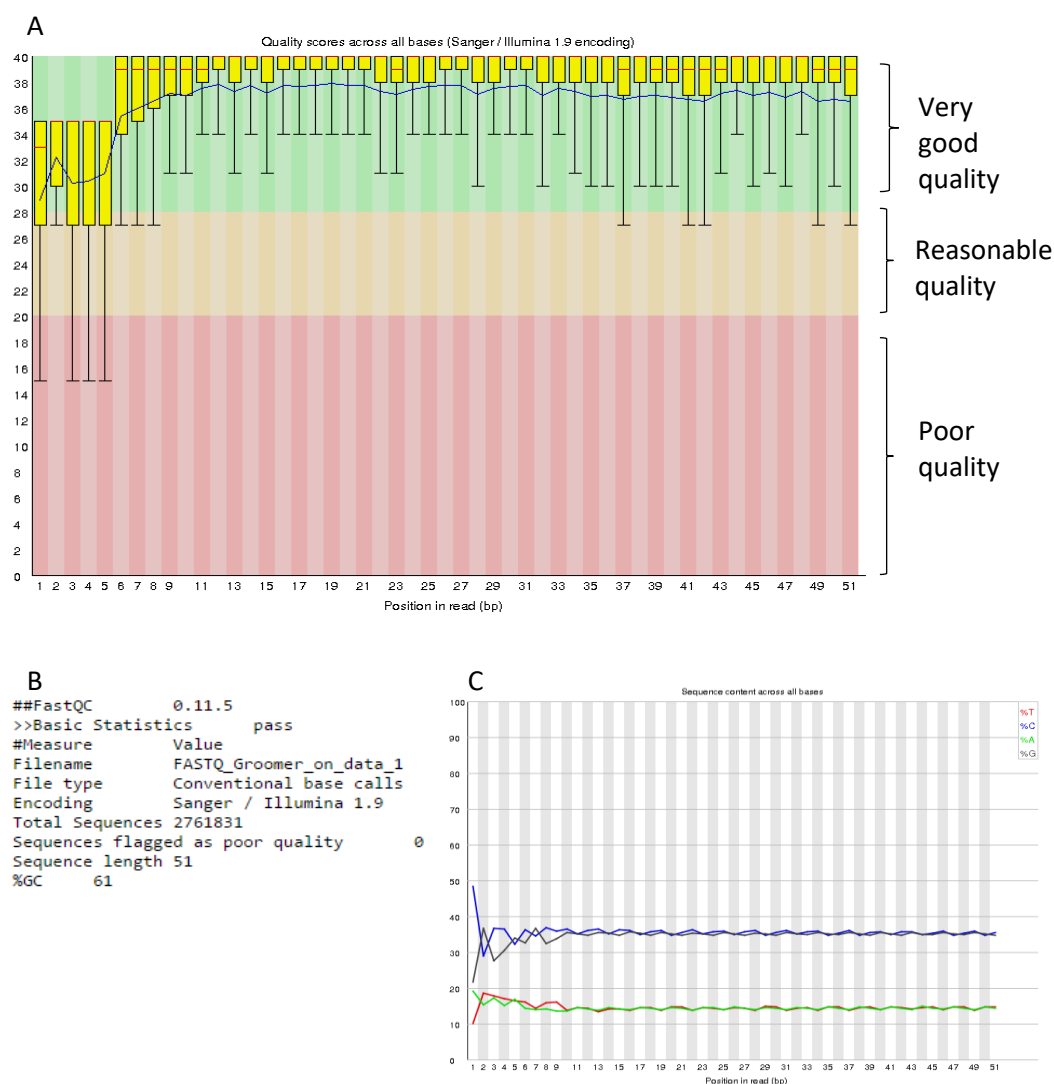


Figure 5.7. Quality control of FASTQ files. (A) The quality score across all the bases, shows that this sample has passed the QC, the first 6-10 bases have been trimmed to ensure that only the ‘very good quality’ bases are included in the analysis. (B) The basic statistics of a sample that has a poor GC%. (C) Sequence content across all of the bases provides an indication of the number of bp require trimming.

The FASTQ files were trimmed from the first 10 bases using the Trim sequences tool (Galaxy Tool Version 1.0.0) to remove the poor sequence quality at the 5’. The FASTQ files were aligned to the *S. coelicolor* A3 (2) genome sequence (NCBI reference sequence: NC_003888.3) with Bowtie for Illumina (Galaxy Tool Version 1.1.2, using the settings `-q -p 6 -S -n 2 -e 70 -l 28 -maxbts 125 -k 1 --un`).

Sample	Mapped Reads	Unmapped Reads	Total Reads	% Mapped
HrdB ^{WT1}	21641143	461501	22102644	97.9%
HrdB ^{WT2}	23023728	940501	23964229	96.1%
HrdB ^{WT3}	18435590	426427	18862017	97.7%
HrdB ^{4xR1}	19167262	336947	19504209	98.3%
HrdB ^{4xR2}	25160294	575662	25735956	97.8%
HrdB ^{4xR3}	21006305	355820	21362125	98.3%
HrdB ^{WT3} IPC	21416330	161127	21577457	99.3%
HrdB ^{4xR3} IPC	21319715	171094	21490809	99.2%

Table 5.2. Percentage of reads mapped to the *S. coelicolor* genome. The number of mapped and unmapped read counts were obtained with IdxStats (Galaxy Tool Version 2.0) after running the alignment to the *S. coelicolor* genome using Bowtie (version 1.1.2).

The percentage of mapped reads were obtained using the IdsStats (Galaxy Tool Version 2.0). This revealed good alignment results to the *S. coelicolor* genome, averaging 98% (Table 5.2). The Bowtie Sequence Alignment Map (SAM) file was then converted to an input BAM file using SAM-to-SAM (Galaxy Tool Version 2.0). This BAM file is then sorted by chromosomal co-ordinates using the Sort BAM dataset (Galaxy Tool Version, 2.0). Finally, the BAM files were uploaded to the Galaxy deepTools server (<http://deeptools.ie-freiburg.mpg.de>) and bigWig files were generated using bamCoverage Tool (Galaxy Tool Version 2.5.0.0). The samples were then normalized using the Normalisation method: Normalized to fragments (reads) per kilobase per million (RPKM). These files were then uploaded to the Integrated Genome Browser (IGB; Version 8.3.4) to visualize the normalized bigWig histograms against the annotated *S. coelicolor* A3 (2) genome sequence (Fig 5.8).

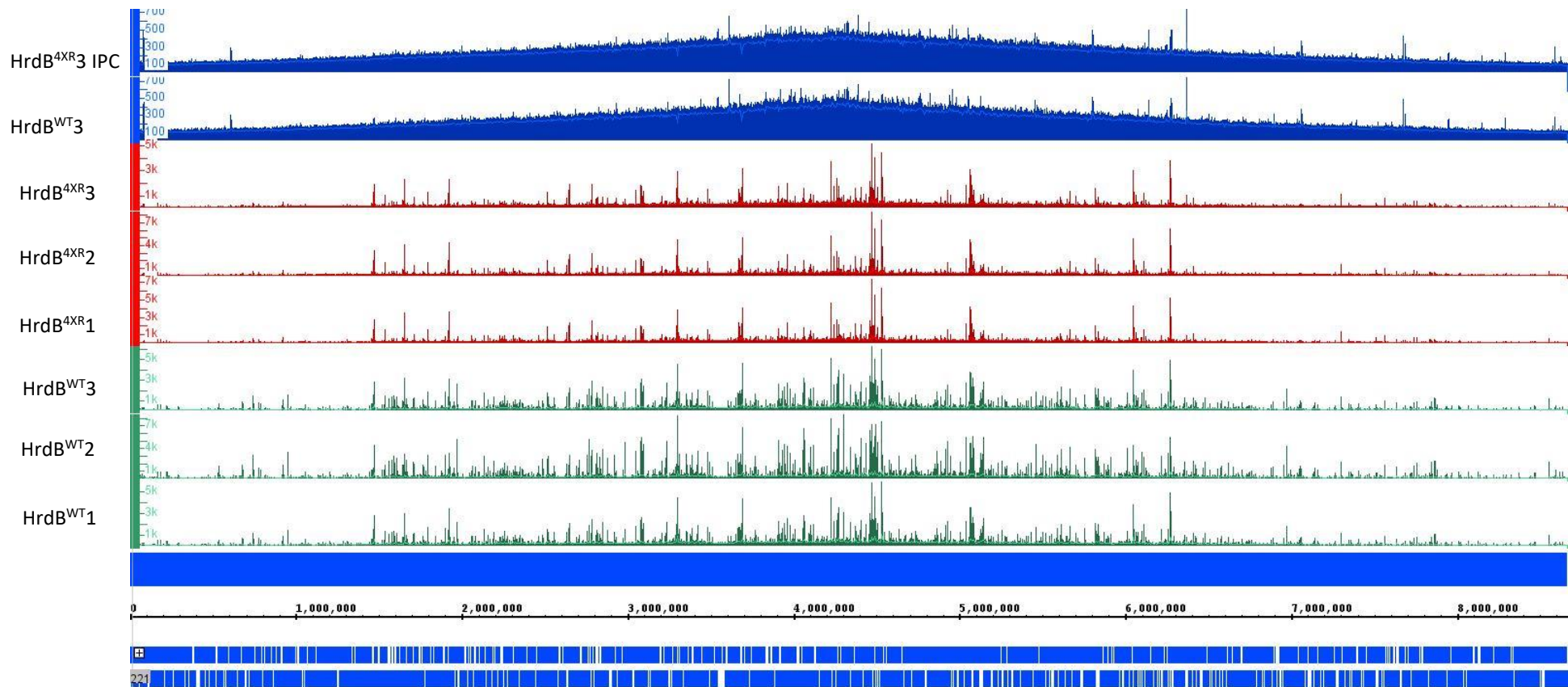


Figure 5.8. bigWig Histogram file visualised using the Integrated Genome Browser. ChIP-seq analysis of the σ^{HrdB} association with RNAP in the absence of RbpA. The bigWig files were aligned to the *S. coelicolor* A3 (2) genome, input controls are in blue (IPC), HrdB^{WT} in blue and HrdB^{4XR} red (immunoprecipitated DNA).

5.2.2.4.2 Analysis of the HrdB^{WT} and HrdB^{4xR} ChIP-sequencing

A previous unpublished ChIP-seq study using a polyclonal antibody mapped genome-wide HrdB binding sites in the presence and absence of rifampicin (which traps RNAP at promoters; A. Tabib-Salazar and M. Paget, personal communication). Therefore, the current set of data were visually compared with this study to ensure that the anti-FLAG antibody was enriching the same promoter DNA. Inspection of two rRNA operons revealed a similar pattern of peaks to that obtained with an anti-HrdB polyclonal and anti-FLAG antibody, although there were less normalised reads, illustrated by the scale bar (Fig 5.9). The lower read count relative to background levels could be caused by many factors including competition between the FLAG-tagged σ^{HrdB} proteins and the native σ^{HrdB} , or deleterious effects caused by the FLAG-tag on holoenzyme formation or promoter binding.

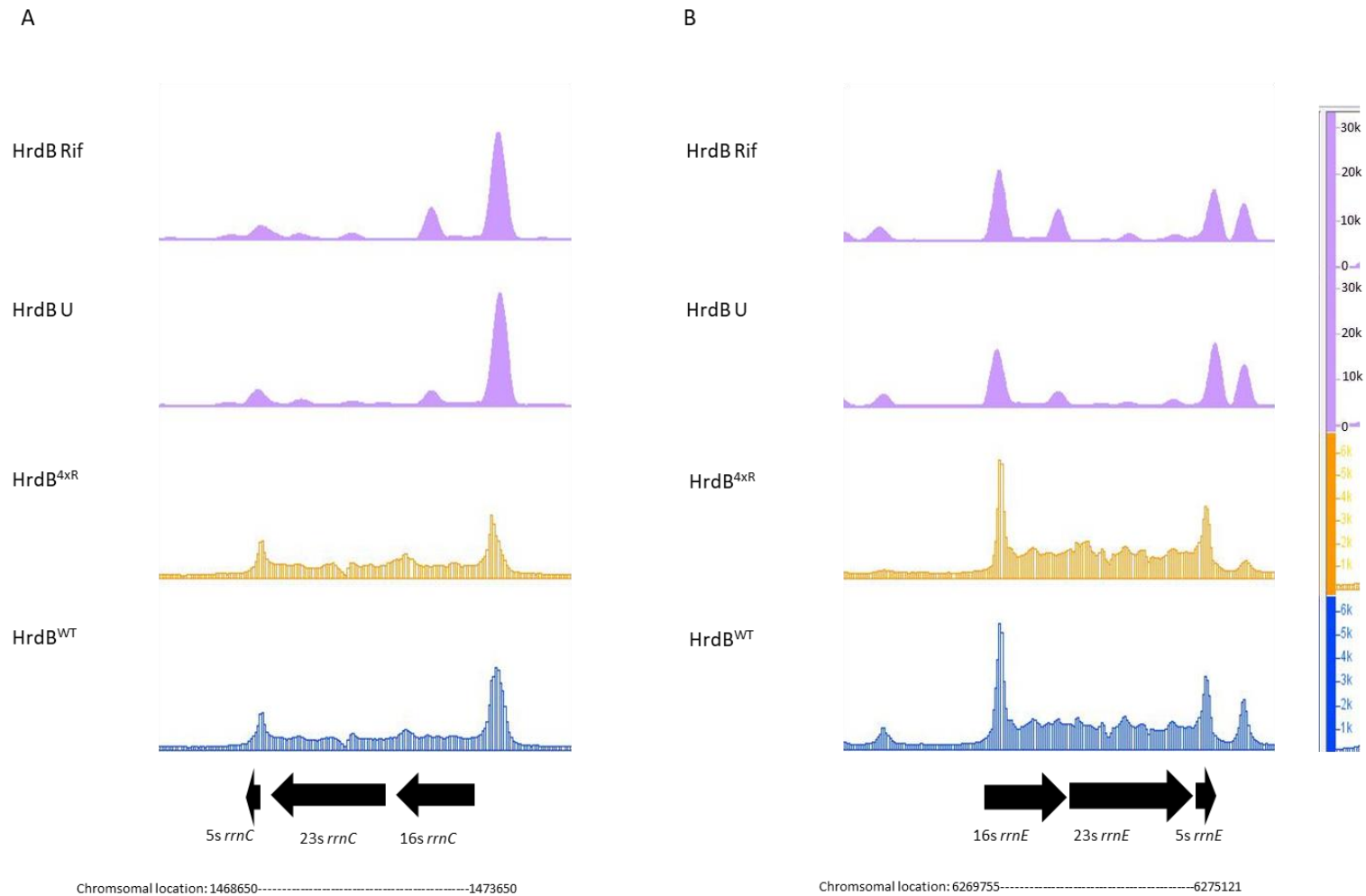


Figure 5.9. Validation of ChIP-seq for this study. (A) Comparing the level of binding of HrdB at the *rrnC* operon and (B) *rrnE* operon. HrdB U (lilac) and HrdB Rif (lilac) are ChIP-seq datasets from Aline Tabib-Salazar, U represents untreated and Rif represents HrdB binding post Rifampicin treatment. HrdB^{WT} (blue) and HrdB^{4xR} (yellow) are comparing the level of HrdB localisation at promoters, HrdB^{4xR} cannot bind RbpA.

The RNA-seq in Results chapter I: 3.1.8.1 demonstrated that the S401 (pRT802), empty vector control which does not express *rbpA*, caused global downregulation of growth related genes. Therefore, as an initial step the level of binding of the HrdB^{4xR} mutant at the highly active tRNA promoters was investigated. For this purpose, 21 5'-mapped promoters for tRNA genes/operons were chosen. The enrichment of HrdB at each tRNA promoter was quantified using the NGS FASTQ files, this was achieved by uploading the files into Galaxy and searching the reads from the HrdB^{WT} and HrdB^{4xR} samples for 20 bp sequence upstream of the +1 transcription start site of the tRNA promoter using the 'Text Manipulation' function and 'Search in textfiles'. These counts were normalised by dividing by the total counts of three control sequences for each sample. The control sequences were located within the long gene, SCO5281 that were not associated with HrdB peaks and therefore represent background reads (Fig 5.10). Almost all of the tRNA gene promoters investigated showed reduced HrdB^{4xR} binding at the same promoters as HrdB^{WT}.

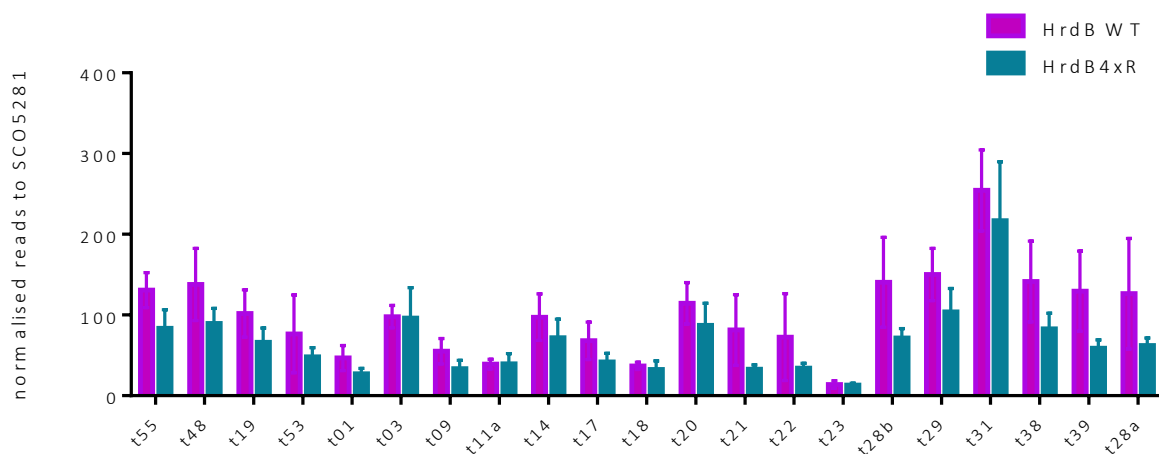


Figure 5.10. The dependency of tRNA genes on RbpA using ChIP-seq. The 20 bp upstream sequences from the tRNA +1 was used to search the reads for each sample using the 'Search in textfiles' (Galaxy Tool version 1.1.0) in Text manipulation (Jeong et al., 2016). The number of reads that produced a hit for the 20 bp sequence from tRNA genes in the HrdB^{WT} and HrdB^{4xR} samples were quantified and divided by the number of the reads for SCO5281. This data is compiled from three biological replicates and error bars represent mean with standard deviation. HrdB^{WT} = purple and HrdB^{4xR} = teal.

Next the annotated TSS were aligned for each of the tRNA gene promoters (Fig. 5.11), in an attempt to identify conservation of particular promoter elements (Jeong et al., 2016). The promoter alignments revealed that the -10 element were well conserved to the consensus sequence, whilst the -35 element was not recognisable. The -35 promoter element is rich in thymine residues, this region is important for binding to the α CTD (Estrem et al., 1999).

SCot01	GTCGGCAAGCGGTT <u>TGATGG</u> AAACGCGGGGCGCCCTGTATTGTTTCAGTGCG
SCot03	TGCACAAACGCGTTT <u>TTGTG</u> CTGGCCCCGGAATCCGCTAGAGTTTCATCACG
SCot09	GCAGGCGGTACTGC <u>TGACAA</u> GAGCGTACCT-CTGCGATAAGGTGGGGGC-A
SCot11	CTGCCCTCGAGACC <u>CGGGAA</u> CCGCGCGGCAC-CTGCGGAGTCCCAAGG-A
SCot14	AGCGGAAACGATTT <u>GGTGT</u> TGTGCCCGGTTCGACCGTGTAATGTTGGCGTCG
SCot17	GGACCCGAGGGT <u>GTGCGT</u> CGAGGGGGGAC-CCCGCTACGATGCGTGC-A
SCot18	AGACGCCAAAAGCG <u>ATGCAC</u> GATTTCGTAAGTCCCTGCGATTATCGGGATCG
SCot19	TCCGGAAGCCGGGCGCGAGCACCCCTTCGGACCCTGTACACTTTTTCTCA
SCot20	TGCGGATTTTCGATT <u>TTTCCT</u> CCCCCGGACCAGGATGCTAAGGTCTACGACG
SCot21	CGGCTGCGCCGGGCGCGGAAGCTGAACCGGAGCGCGGCACCGGTGCTG
SCot22	ATTCCCCTCACACTCCCGGTCCGCAGGTCCGCGCGGTACGGTTCGGTACA
SCot23	TCAAGGACTTGGGGATCTGACGGCGCTGCCCGACCATGTAGCATGGCCGG-G
SCot28a	TACCTGCTGACCGGCACCCGCGAGGACGGCAAGCGCTGCTCGTAAC-C
SCot28b	CCTCCGGGGCGGGTTCGGAGCACCCCGGGAACGTGTAGACTTGCCGACG
SCot29	CCGATAAACCGGATTTCTGCTCCCGGCCACCGGTGGCTAAAGTAAACGA-C
SCot31	GGGGGAAAGGGATGTGAATCGACCCCTGTCAGCGTGCTAATGTTCTCGT-G
SCot38	CCCCAGAACGAGTTCCGAAGCACTCCTCGGAGTCAATGTAATGTTCTCCGCG
SCot39	CACGCGGCGGTGCGGCTCGCGAGGGAGCGCGGTTGGTATAGTTGGCTGTCTG
SCot48	CGATCAATCGGGTGCGGAAGCACCCCTCCGCATCAGGTATTGTTTACGTCTG
Scot53	CGTGAGCAGGGATTGCGCTCACGTCTTCATCGTGACGTAGTGTTCATTCA
Scot55	CATGACCACCCCGGCGGAGCTGGTAGTGTTCTACCGCTCGCCCGGAGCAAC

TTGaCA N17 TAgaaT +1

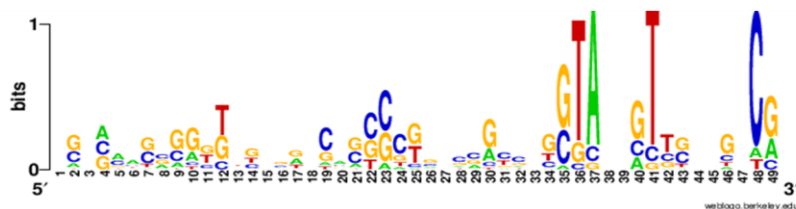


Figure 5.11. tRNA gene promoter alignments. The TSS were identified from Jeong et al., 2016, bases in grey represent position -1 from the TSS. The -10 and -35 promoter elements are underlined and bases in red are the extended -10 element. Motif conservation created using WEBLOGO (version 2.8.2), position 1 represents -49 and position 49 at the 3' presents +1 of the promoter.

Initial attempts to identify peaks for HrdB^{WT} and HrdB^{4xR} using the input controls in MACS2 and then identify differences bioinformatically using bdgdiff (version 2.1.1.20160309.0) failed because the differences were small. Additionally, MACS2 is ideal for identifying the presence/absence of peaks rather than less/more binding at the same locus. Therefore, HrdB^{WT} and HrdB^{4xR} peaks were compared using MACS2 directly without the use of input controls. Each biological replicate comparison produced a list of genes, which were pooled together, and a 1.5-fold enrichment filter was applied, resulting in a list of 36 genes (Table 5.3). This list of 36 genes represented less binding for the HrdB^{4xR} samples vs HrdB^{WT}. The list of genes also showed that the vast majority of differences between HrdB^{WT} and HrdB^{4xR} exist at the periphery of the chromosome (Fig 5.8).

Gene ID	start	end	Fold enrichment	-log10 (q-value)	Function
SCO6032/SCO6033	6622487	6622846	1.55291	8.05193	Hydrolase/ hypothetical protein
SCO6057/SCO6058	6649976	6650200	1.52897	8.05193	ATP/GTP-binding integral membrane protein/hypothetical protein
SCO6960	7726906	7727168	1.52833	10.81185	Hypothetical protein
SCO6468	7157227	7157558	1.52619	10.81185	phosphatidylserine decarboxylase
SCO6864	7635024	7635264	1.52436	10.7076	hypothetical protein
SCO7815/SCO7814	8634726	8634940	1.52436	10.7076	TetR family transcriptional regulator

SCO1018/SCO1019	1073953	1074192	1.52039	10.55826	glutamate racemase/ integral membrane protein
rrnc	1471828	1472038	1.51642	10.4136	Ribosomal protein
SCO6374	7036266	7036543	1.51512	10.61922	sugar transferase
SCO0297	291595	291853	1.51245	10.26669	secreted protein
SCO1302	1379640	1379868	1.51245	10.26669	hypothetical protein
SCO7552/SCO7553	8374464	8374696	1.51245	10.26669	TetR family transcriptional regulator/ oxidoreductase
SCO0089/SCO0088	75815	76029	1.50054	9.82452	Transcriptional regulator/ secreted beta-lactamase
SCO1242	1314390	1314647	1.50054	9.82452	DNA-binding protein
SCO1544/SCO1543	1652478	1652848	1.50054	9.97343	Hypothetical proteins
SCO6863	7632947	7633194	1.50054	9.82452	hypothetical protein
SCO0504	537420	537716	1.64775	14.07565	DEAD-box RNA- helicase
SCO0748	789707	789962	1.57873	11.53981	hydrolase
SCO0636	676905	677192	1.56003	11.53052	ABC transporter ATP- binding protein
SCO1510/SCO1509	1614012	1614270	1.55056	10.75063	peptidyl-prolyl cis- trans isomerase/ hydrolase
SCO1656	1772625	1772903	1.53856	10.49098	hydrolase

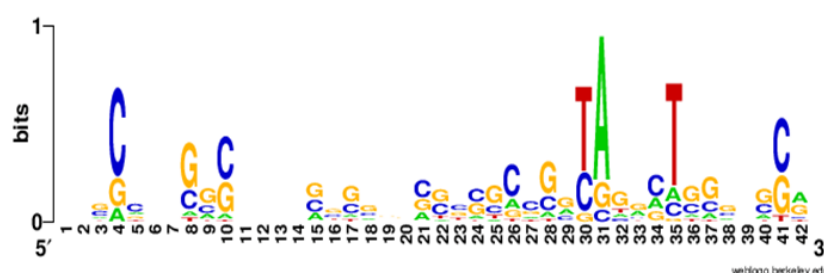
SCO6391	7056806	7057424	1.53176	10.6936	IS110 transposase/integrase
SCO7203	8006039	8006283	1.53129	9.74129	integral membrane protein
SCO7253	8065015	8065237	1.53129	9.74129	hypothetical protein
SCO7093	7881725	7881978	1.51737	9.42179	transcriptional regulator
SCO1365/SCO1364	1442621	1442862	1.51403	9.12645	Hypothetical protein
SCO1491	1594068	1594473	1.51108	10.18576	elongation factor P
SCO6336/SCO6335	6995657	6995907	1.50972	8.98058	Hypothetical protein
SCO0233	223713	223920	1.50541	8.82848	DNA-binding protein
SCO7318	8125286	8125498	1.50541	8.82848	hypothetical protein
SCO6294/SCO6295	6954253	6954566	1.50339	9.88262	GntR family regulatory protein/ ABC transporter ATP- binding protein
SCO6488	7157250	7157488	1.50041	8.82848	acyl-peptide hydrolase

Table. 5.3. List of enrichment sites identified as significantly lower for HrdB^{4xR} compared to HrdB^{WT} using ChIP-seq. MACS2 provides a list of enrichment at chromosome locations, which were then viewed using IGB (version 9.0.0). The gene ID column contains two gene IDs because the enrichment could be present for either strand. Start and end refer to the chromosomal location.

The promoters from the genes in Table 5.3 were aligned using TSS annotations to identify unique elements of the promoters of genes which all demonstrate less binding of HrdB in the HrdB^{4xR} compared to HrdB^{WT} (Fig 5.12). RbpA has been shown to activate *in vitro* transcription

of promoters which lack a consensus -35 promoter element so it was important to check the quality of the -10 vs -35 promoter sequences (Hubin et al., 2017a; Vishwakarma et al., 2018). The alignment of the promoters revealed almost perfect conservation of the -10-consensus (TAgaaT) sequence, whilst the -35 (TTGaCA) promoter sequence was poorly conserved. This result suggests that RbpA is important for binding to promoters with poor -35 consensus sequences.

TTGaCA	N17-20	TAgaaT
-35		-10



184

5.4 The genes that require RbpA in *S. coelicolor*

RbpA is present at all σ^{HrdB} -dependent promoters *in vivo*, although it is not clear which promoters actually require RbpA for expression. RbpA appears to stimulate most promoters *in vitro* (Tabib-Salazar et al., 2013), although other cellular factors, such as activators, are missing from these experiments, and so the actual *in vivo* dependency on RbpA is unclear. The question of which genes are dependent on RbpA was partially addressed in Results chapter I, where the role of the basic linker was investigated. However, considering that the basic linker is not essential, RbpA could play additional roles at some promoters. To try to address this we used RNA-sequencing to compare the transcriptome of the S401 (pRT802) and S401 (pRT802::*rbpA*^{WT}) strains under conditions where the thiostrepton-induced *rbpA* gene in S401 was not expressed. Following initial characterisation of the data, the focus was on genes that were up-regulated in S401 (pRT802::*rbpA*^{3KRA}) and differential gene expression between S401 (pRT802) and S401 (pRT802::*rbpA*^{WT}). For information on the library preparation, quality control data and the level of adapter contamination and trimming, refer to Results I; sections 3.1.6.1, 3.1.6.2, 3.1.7.1 and 3.1.7.2.

5.4.1 Analysis of the differential gene expression between S401 (pRT802) and S401 (pRT802::*rbpA*^{WT})

A differential gene expression list was generated comparing S401 (pRT802) and S401 (pRT802::*rbpA*^{WT}), 3562 genes were significantly different, with 1352 genes up-regulated in S401 (pRT802) and 2210 genes down-regulated. The samples were loaded into Rstudio (version 1.0.153) using the CummeRbund package (version 2.0.0) to visualise the samples (Fig 5.13A and Fig 5.13B) biological replicate FPKM values. The FPKM values for the three biological replicates for S401 (pRT802) and S401 (pRT802::*rbpA*^{WT}) were demonstrated using box plots and dendograms. The dendograms showed that the biological replicates cluster together based on the genetic background (Fig 5.13A), whilst the box plots revealed the conservation of the biological replicates (Fig 5.13B).

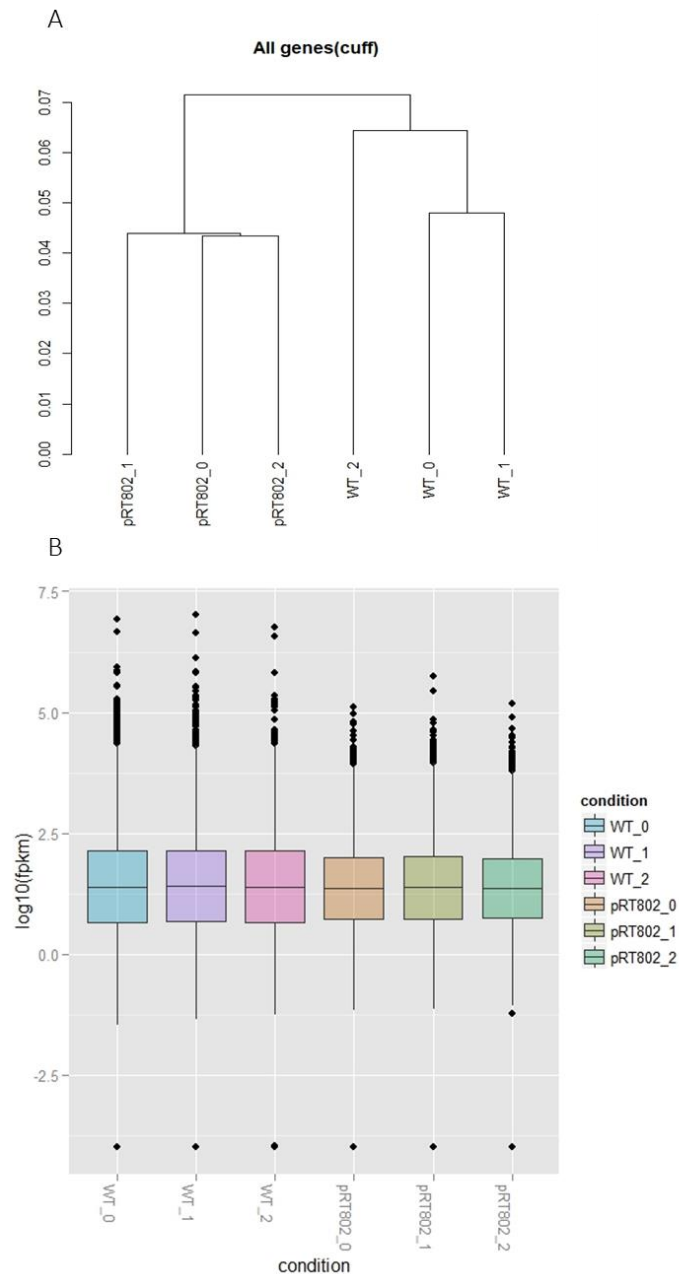


Figure 5.13. Quality of the replicate RNA-Seq samples in *S. coelicolor*. (A) Dendrogram demonstrates the hierarchical clustering of the three-biological replicate. (B) Density plot of the distribution of FPKM values for all three-biological replicates. All the quality control measures have been created using the cummeRbund package (version 2.0.0) in Rstudio (version 1.0.153).

5.4.2 Analysis of genes up-regulated in S401 (pRT802)

To analyse the up-regulated genes in S401 (pRT802), the differential gene list was stringently filtered, firstly by $>\log_2=1$ difference in expression level and secondly by filtering out the S401 (pRT802) FPKM values <20 . The resultant list contained 200 genes and was sorted by the chromosome location to identify clusters of genes.

5.4.2.1 Genes induced by cell wall antibiotics are induced in S401 (pRT802)

S401 (pRT802) has shown a clear growth-related phenotype, the absence of *rbpA* expression in this strain may be responsible for the induction of a multitude of stresses. In line with this, the bacterial cell wall is a good target for antibiotics as it is crucial for growth. Therefore this section focuses on the changes in expression of genes that are upregulated in response to cell wall antibiotics. A study by the Hong lab (2011) investigated the genes up-regulated by vancomycin, bacitracin and moenomycin which target the cell wall in *S. coelicolor* (Hesketh et al., 2011). This study found 1062 significantly induced genes in response to exponentially growing *S. coelicolor* in response to vancomycin, bacitracin or moenomycin. The list of 244 genes induced by all three cell wall targeting drugs were compared to the RNA-seq data from this study which revealed a match of 37 genes. This list of 37 up-regulated genes included *hrdD*, which is controlled by SigE. This raised the possibility that the SigE regulon is induced in the S401 (pRT802) background. Upon investigations of the SigE regulon, 22 of the SigE-dependant genes were identified and matched to the 37 genes cell wall antibiotic targeting genes. SigE (SCO3356) is an extra-cellular sigma factor (ECF) which is activated during cell envelope stress (Paget et al., 2002). The only known SigE targets to date was σ^{hrdD} and the *cwgA* operon (Paget et al., 1999). Next the RNA-seq data comparing S401 (pRT802) and S401 (pRT802::*rbpA*^{WT}) revealed 28 possible SigE targets in total. However, the levels of *sigE* expression in *S. coelicolor* S401 (pRT802) and S401 (pRT802::*rbpA*^{WT}) were unchanged, whilst SigE was down-regulated in S401 (pRT802::*rbpA*^{3KRA}) by 1.88-fold. The annotated TSS for 20 up-regulated genes of the 22 potential SigE targets were aligned to identify the conservation of the -10 and -35 promoter elements (Table 5.4; 5.14). The promoter alignments showed conservation of the -35 (GgAAC) and -10 (TCtpy) promoter elements which were recognised by SigE (Kang et al., 1997).

Gene ID	Function	Promoter sequence
SCO1875	secreted penicillin binding protein	GAAATCGGGGCACACG <u>GGAAC</u> GACCGCCGCCCGCCCGTTCG <u>TCCTGT</u> -C
SCO3089	Putative ABC transporter ATP-binding protein	GGAACAGCGGCTGACT <u>ACAGG</u> GCGCCGTTCGACCGGGACGAT <u>TTGACAAC</u>
SCO3956	Putative ABC transporter ATP-binding protein	GGCTCTGCGGTCTGACG <u>ACCGC</u> AGTGGCGGACGTCCTACTG <u>TCATT</u> GAC
SCO3761	hypothetical protein	CGACGGGGTCG-TCGG <u>CGAAC</u> TC-CGGGTTCGCTTCGCCG <u>TTGCGG</u> -C
SCO2939	hypothetical protein	CGCCCGCGCGGTGAG <u>GCAAC</u> GAGTGCCTCCCCCAC-GCG <u>TCCTAC</u> -T
SCO2892	secreted protein	CTTCCTGACCGACTCCC <u>GGAAC</u> GGAACA-CAAGTCCCAGCG <u>TCTGT</u> -C
SCO3397	integral membrane lysyl-tRNA synthetase	CCGTTGATCAGGGGAGT <u>TGAAC</u> CTCT-CCCTCCGAGACACCG <u>TCCTAG</u> -C
SCO1023	hypothetical protein	AAACGACCCGGTAACG <u>GGAAC</u> CCCGT-CCTGGGTCCGCGCGT <u>TTGGG</u> -T
SCO4157	Putative protease	CTGCCGGGCTTCTCAG <u>GGCAG</u> CACTC---AGGCCGCTCTCAG <u>CGCGG</u> --A
SCO1403	Putative membrane protein	GCGCGGGATTTTCG <u>GGAAT</u> CTTGG-GGGCCCTTCGTTCTG <u>TCCTT</u> -G
SCO2582	Conserved hypothetical protein	CTCCTTCGACGTACT <u>GCCCC</u> TCCACGGTATCTCCGCCTAG <u>GCTGGG</u> -G
SCO1867	Putative hydroxylase	CCGCGAGACGG <u>GGAAG</u> ACGTA-CACGCCGGGCACGAT <u>TGTAT</u> -C
SCO1700	Putative membrane protein	GCCCTCGACG <u>TGGAC</u> GCGGC-GCTGACCTGGCGGGT <u>TCGCCG</u> -T
SCO0608	Putative regulatory protein	TTGCGGCGGCACAT <u>ACCAC</u> GAGGG-GCCATACGTGGTAT <u>ACCGA</u> -G
SCO1023	Putative membrane protein	AAACGACCCGGTAACG <u>GGAAC</u> CCCGTCCCTGGGTCCGCGCGT <u>TGGGCT</u> -T
SCO0226	Putative membrane protein	GCTGCCGCCCTCGC <u>GGAAC</u> CGACC-GGCAAGGAAAGTGAT <u>TCAAT</u> G-A
SCO1695	Hypothetical protein	CGGATGCGGTCTTCAG <u>GGAAC</u> CCGTGGAAGCCGCGGGAGT <u>TCTCA</u> G-G
SCO1903	Putative transport associated protein	TGGCCCCCGGGTAAG <u>GGAAC</u> GGTAAAACGCCCTCACTCGT <u>TCACC</u> -A
SCO2665	Hypothetical protein	TACGGCGTGCGATGA <u>AGAAG</u> GACG-CCGACGATCTGGTAC <u>GCCGG</u> G-A
SCO3202	RNA polymerase alternative sigma factor	GACTCCGCGTCCGTG <u>GCAAC</u> CCTC-AGGCGGTACGGGCCG <u>TCTTCA</u> -G

GgAAC

N18-19

Tc_{tpy} +1

N2-6

Table 5.4. Promoter alignments of the SigE targets up-regulated in S401 (pRT802). The transcriptional start site data used from Jeong et al (2016) has been used to identify the promoters for the SigE genes. The -10 and -35 elements are separated by 18-19 bases and the TSS +1 position of the genes is shown as +1.

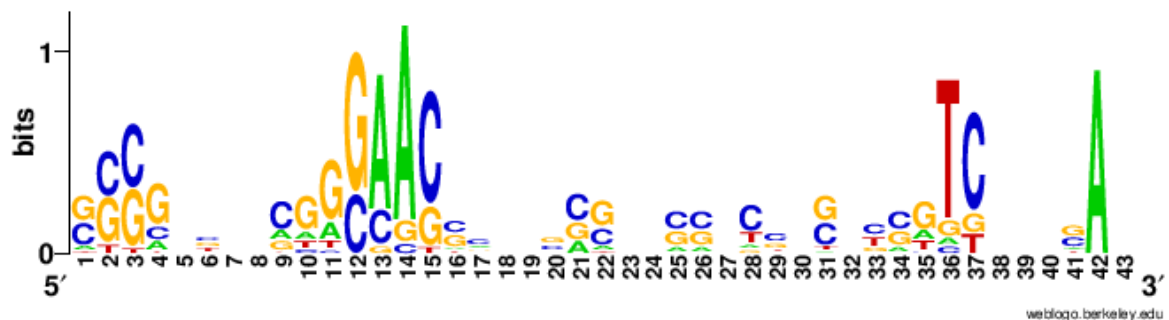


Figure. 5.14 Motif conservation of the SigE targets. The SigE targets motif conservation has been generated using WEBLOGO (version 2.8.2), position 1 represents -43 and position 43 at the 3' presents +1 of the promoter

Interestingly, SCO3089-SCO3090 are ABC transport proteins known to be up-regulated in response to bacitracin treatment and null mutants demonstrate increased sensitivity. The RNA-seq data showed that SCO3089-SCO3090 as induced by 11.6 and 12.8-fold in S401 (pRT802), respectively. Additionally, SCO3397 is the only lysyl-tRNA synthetase out of a possible 27 that is up-regulated by 2.1-fold, it has been also known to play a role in resistance to bacitracin and vancomycin (Hesketh et al., 2011).

5.4.2.2 Increased levels of SigE associate with RNAP in $\Delta rbpA$

The RNA-seq results indicated that the SigE regulon is up-regulated in S401 (pRT802::*rbpA*^{3KRA}), despite the levels of *sigE* remaining unchanged in both samples. It was hypothesised that depletion of RbpA might result in increased accessibility of σ^E to the RNAP. Therefore, the levels of SigE in cell lysate and pull-downs were tested by comparing the *S. coelicolor* J1981 (*rpoC*::his) and S129 (J1981 $\Delta rbpA$) strains. To test this, firstly, the starting number of spores were

normalised between both strains and cultured to OD₄₅₀ 0.8 and the cell lysates were passed through the cobalt protein purification column. The pull-down samples were normalised to J1981 and run on a 10% SDS gel and imaged using LI-COR. The western blot analysis showed the level of SigE in the cell lysates in J1981 and S129 was generally consistent whilst the levels of SigE associating with RNAP in the pull-down assays indicated an increase in SigE association with RNAP in the S129 strain (Fig 5.15). The anti-SigE antibody appeared to bind to several other proteins shown in Fig 5.15A, B. SigE is a 28 kDA protein and this was used to line up the band corresponding to SigE (Kang et al., 1997). This data suggested that the expression levels of *sigE* are unchanged in the absence of RbpA, however the levels of SigE associating with RNAP increases thus resulting in the up-regulation of the SigE regulon.

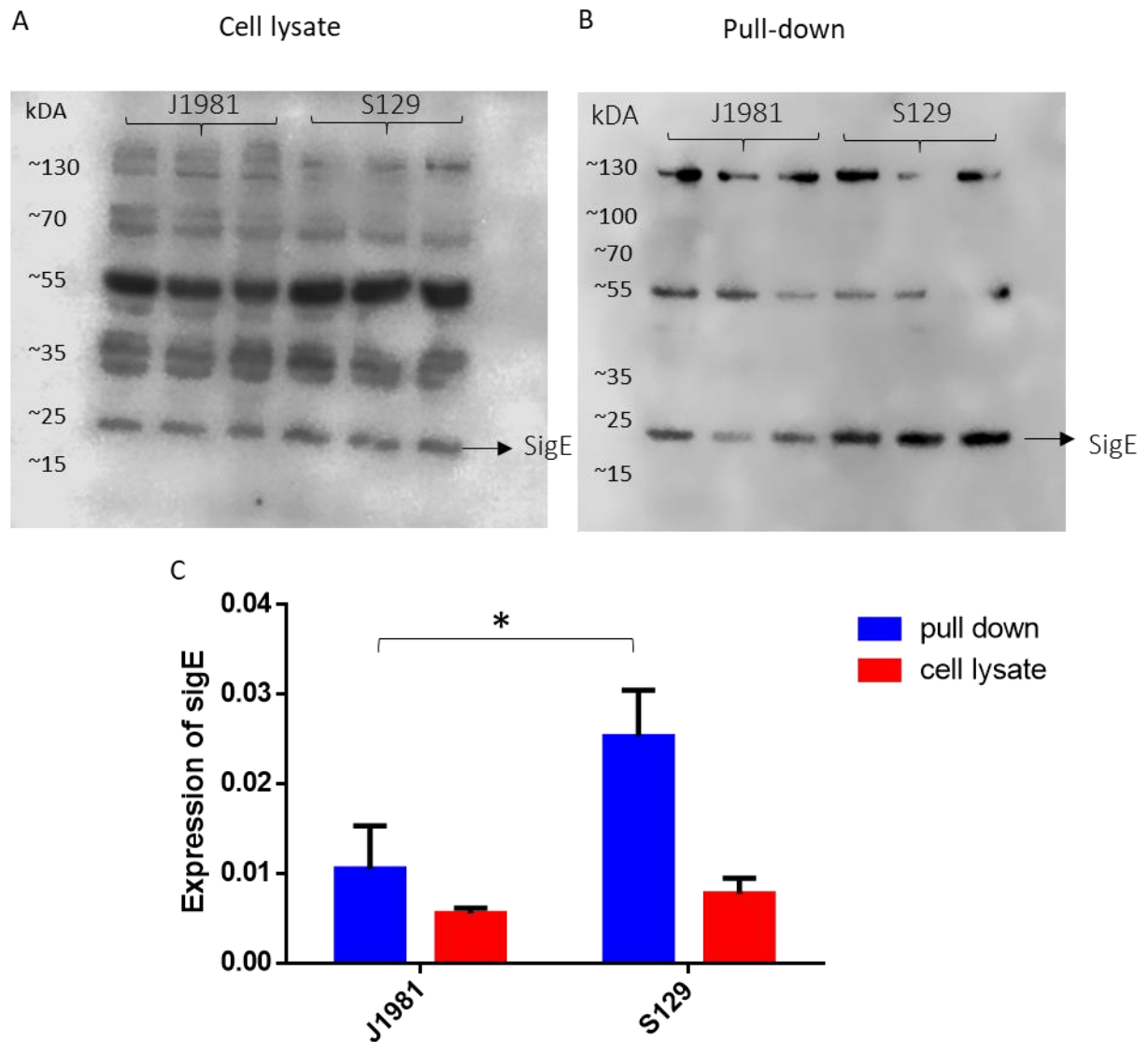


Figure 5.15. Increased association of SigE in the S129 $\Delta rbpA$. Spore stocks for J1981 and S129 strains were normalised and cultured to OD₄₅₀ 0.8, the cell lysates were passed through the cobalt protein purification column following manufacturer's instructions. The pull-down protein concentrations were normalised to J1981 and run on 10% SDS gel. The primary antibody anti-rabbit anti-SigE and the secondary antibody anti-rabbit anti-HRP was used to blot the membrane and visualised using the LI-COR. Multiple bands on the western blots indicate that the anti-sigE Ab is able to detect other proteins (A) Cell lysate (B) Pull-down samples visualised and quantified (C). Unpaired T-test compares the J1981 and S129 pull-downs, p-value 0.0224. This data is compiled from three biological replicates and the error bars denote mean standard deviation.

5.4.3 Upregulation of genes in S401 (pRT802::*rbpA*^{3KRA}) and S401 (pRT802)

The up-regulated genes in S401 (pRT802::*rbpA*^{3KRA}) were not discussed in Results Chapter I, and therefore has been presented in this section. Investigating genes that are both up-regulated in S401 (pRT802::*rbpA*^{3KRA}) and S401 (pRT802) could provide an insight into gene regulation by RbpA. Therefore, the list of 2768 genes that are significantly up-regulated in S401 (pRT802::*rbpA*^{3KRA}) and S401 (pRT802) was stringently filtered. Firstly, the list was filtered by log₂=1 (2 fold change) for both datasets, secondly genes with FKPM <20 for both S401 (pRT802::*rbpA*^{3KRA}) and S401 (pRT802) were removed. This produced a list of 91 genes which were organised based on chromosomal location for the identification of gene clusters.

5.4.3.1 Carotenoid biosynthesis operons, *crtEIVB*, *crtYTU*, *litRQ* and *litSAB* are up-regulated in S401 (pRT802) and S401 (pRT802::*rbpA*^{3KRA})

Carotenogenesis is the production of natural pigments which can provide protection against photo-oxidative damage and aids light harvesting (Takano et al., 2005). The biosynthesis of the yellow-pigmented carotenoids in *S. coelicolor* are light-inducible and the sequenced genome enabled identification of the carotenoid biosynthesis gene cluster (Takano et al., 2005). The carotenoid gene cluster is composed of two convergent operons, *crtEIBV* and *crtYTU* and the cluster is flanked by the light-induced regulatory divergent operons, *litRQ* and *litSAB* (Takano et al., 2005). LitR is part of the MerR transcription family of regulators and LitS is an ECF sigma factor. LitA is a lipoprotein, LitB is a putative anti-sigma factor and LitQ is a possible oxireductase. Transcription at the promoters proceeding *crtE*, *crtY*, *litR* and *litS* are activated upon light illumination (Takano et al., 2005). Sigma factor LitS drives the activation of the *crt* biosynthesis gene cluster upon light activation (Takano et al., 2005). The *crtEIVB* and *litS* deletion mutants failed to produce the carotenoid yellow-pigment (Takano et al., 2005). The RNA-seq from this study showed an up-regulation of the convergent operons *crtEIBV* and *crtYTU* in S401 (pRT802) by 16 and 15-fold and in S401 (pRT802::*rbpA*^{3KRA}) by 14 and 12.1-fold, respectively. Additionally, *litQ*, *litR*, *litA* and *litB* are up-regulated 3.6, 1.3, 3.8 and 8.2-fold in S401 (pRT802) respectively and *litB* is up-regulated 3.5-fold in S401 (pRT802::*rbpA*^{3KRA}). Interestingly the expression levels of sigma factor, *litS* remains unchanged between S401 (pRT802::*rbpA*^{WT}), S401 (pRT802) and S401 (pRT802::*rbpA*^{3KRA}).

5.4.4 Down-regulation of genes in S401 (pRT802)

Results chapter I;3.1.8.1 focused on the growth-related genes which were all down-regulated possibly as a direct result of the RbpA null mutant, RbpA^{3KRA} or because the strain simply failed to grow. This section focuses on the genes that are significantly down-regulated in S401 (pRT802) but not in S401 (pRT802::*rbpA*^{3KRA}). There was a list of 361 down-regulated genes in S401 (pRT802) which were stringently filtered by applying a FPKM value <20 for S401 (pRT802) which reduced the list to 210 genes and secondly applying a log2=-1 fold cut-off, resulting in a final list of 20 genes.

S. coelicolor encodes two glutamine synthetases, *glnA* which is regulated by GlnR and GlnII (Tiffert et al., 2008). GlnR is a global nitrogen response regulator and binds upstream of the *amtB-glnK-glnD* operon and recognises the GGTCAC-N(5)-CGAAC-N(19-23)-TSS motif of the promoters (Fink, 2002). *amtB* encodes a putative ammonium transporter, *glnK* is a signal peptide and *glnD* encodes for an adenylyl transferase. A study mined the genome sequence for *S. coelicolor* for the GlnR binding motif in the promoter sequences of genes found upstream of the TSS and were confirmed using electrophoretic mobility shift essays (EMSAs) for an extensive identification of the GlnR regulon (Tiffert et al., 2008). The putative secreted protein, SCO0888 was identified as a member of the GlnR regulon which is down-regulated in S401 (pRT802) by 2-fold. Some possible explanations include that, RbpA may regulate the GlnR promoter, the absence of RbpA causes global shut-down in transcription which includes the GlnR regulon.

5.5 Discussion

5.5.1 The SID RbpA^{M85} displays a clear phenotype

The RbpA^{M85} residue is very well conserved across the Actinobacteria and is located in the RbpA-SID which is important for σ^{HrdB} interactions. In this study, the RbpA^{M85} was substituted with leucine, alanine and glutamate. All three substitutions in a $\Delta rbpA$ S129 background displayed minimal changes whilst introduction of the same constructs into an $\Delta rbpA \Delta rbpB$ genetic background resulted in RbpA^{M85E} and RbpA^{M85A} displaying a small colony phenotype and production of ACT. The RbpA^{M85L} does not display a phenotype despite its larger R group

compared to RbpA^{M85A}. Previous studies tested the *in vitro* transcription activation ability of the *M. tuberculosis* RbpA^{M84A}, which has shown partial activation. The corresponding RbpA^{M85A} in *S. coelicolor* was tested in a $\Delta rbpA$ S129 background which produces a small colony phenotype (Hubin et al., 2015). A crystal structure of the *M. tuberculosis* RbpA-SID- σ^A_2 complex, positioned the RbpA^{M84} residue to bind to the σ^A -residue K334 which interacted with -12 template DNA (Hubin et al., 2015). Together this evidence suggests that the RbpA^{M85} in *S. coelicolor* is critical for normal *rbpA* function in vivo. To further characterise the importance of the RbpA^{M85}, growth curves are required to establish the effect of substituting M85 with a negative charged amino acid and those with a larger R group on viability.

5.5.2 σ^{HrdB} unable to bind to RbpA is still able to associate with RNAP and promoters

Previous work by our group identified several amino acids in σ^A_2 positioned to interact with RbpA-SID (Hubin et al., 2015). These amino acids in σ^A_2 were substituted with the corresponding σ^{HrdC_2} residues which rendered RbpA binding in BACTH assays in *M. tuberculosis* (Hubin et al., 2015). Therefore, this study investigated the effects of σ^{HrdB} binding to RNAP by substituting the amino acids important for RbpA binding with the corresponding σ^{HrdC_2} in *S. coelicolor*. The engineered HrdB^{4xR} mutant was analysed for its association with RNAP using pull-down western blot analysis upon diamide treatment. Diamide activates alternative sigma factors to respond to the oxidative stress and therefore may encourage greater competition between σ^{HrdB} and other sigma factors for RNAP. The western blot analysis showed that both HrdB^{WT} and HrdB^{4xR} continue to associate with RNAP. The RNA-seq data for S401 (pRT802) had shown global down-regulation of growth related genes therefore the ChIP-seq data was analysed for growth related genes. The highly active tRNA promoters confirmed using ChIP-sequencing, that the amount of HrdB^{4xR} binding to tRNA genes was only reduced but not abolished in the absence of RbpA, compared to HrdB^{WT}. Promoter alignments of the tRNA genes demonstrates a well conserved -10 whilst the -35 element is poorly conserved. This eludes to the possible necessity of RbpA at promoters with a poor consensus -35 promoter element. The ChIP-analysis involved producing a peak calling list of genes that have less HrdB^{4xR} association compare to HrdB^{WT} at promoters. The promoter alignments of these genes that

were identified again revealed a poorly conserved -35 consensus sequence compared to the -10-promoter element. Other studies use the *vapB10* gene which is annotated as a possible antitoxin in *M. tuberculosis* and is highly expressed, consists of a promoter that has a 'poor' -35 element and has been shown to be dependent on RbpA (Hubin et al., 2015). The *rrnAP3* is a ribosomal RNA gene and its *AP3* promoter also in *M. tuberculosis* consists of a 'good' -35 element and is not dependent on RbpA (Hubin et al., 2015). One study tested the dependence of *vapB*, *AP3* and *AP3^{anti-35}* (-35 from *vapB* swapped with the *AP3* -35 element) promoters on RbpA and CarD and found that both transcription factors together increased transcription activity of *vapB* and *AP3^{anti-35}* more so than *AP3* (Hubin et al., 2017). In addition to this, it does not necessary mean that the HrdB^{4xR} is functional, it is possible that its ability to dock at promoters is unaffected, but it is unable to activate transcription.

5.5.3 Identification of targets dependent on RbpA

RbpA has been shown to bind to group I and some group II sigma factors (Hubin et al., 2015; Tabib-Salazar et al., 2013). In addition to this, some promoters such as the probable antitoxin gene, *vapB10* in *M. tuberculosis* has been shown to be dependent on RbpA *in vitro* (Hubin et al., 2015, 2017). To date, no other targets that are dependent on RbpA *in vivo* have been identified. Therefore, an RNA-sequencing experiment was implemented by comparing RbpA^{WT} to $\Delta rbpA$ (pRT802, empty plasmid). The differential gene expression between the two strains revealed 3562 genes that were differentially expressed, of these 1352 genes were up-regulated and 2210 were down-regulated. However, these lists were too large to enable any significant analysis therefore stringent filters were applied.

5.5.3.1 Up-regulation of genes in the absence of *rbpA*

The RNA-seq data revealed a match to the up-regulation of 37 genes induced by cell wall antibiotics, of these 22 belonged to the SigE regulon (Hesketh et al., 2011). This list of SigE included the already characterised HrdD and DagA (Paget et al., 1999). The promoter alignments of the SigE targets revealed conservation of the -35-consensus sequence (GgAAC) and -10 (TCtpy). The expression levels of SigE are unchanged in S401 (pRT802::*rbpA^{WT}*) and

S401 (pRT802) therefore western blot analysis comparing the levels of SigE using cell lysate and pull-downs from a RbpA^{WT} and $\Delta rbpA$ strain were used to identify the levels of SigE associating with RNAP. The western blot analysis confirmed that the levels of SigE remain the same in the cell lysate however, more SigE associates with RNAP in the $\Delta rbpA$ strain. This is possibly because in the absence of RbpA, HrdB forms unstable RNAP- σ^{HrdB} holoenzyme and can easily be disassociated, therefore the competition between σ^E and σ^{HrdB} increase for RNAP (Mauri & Klumpp, 2014). Lastly, the genes that are up-regulated in both S401 (pRT802::*rbpA*^{3KRA}) and S401 (pRT802) were analysed after applying stringent filters. This highlighted the up-regulation of the light-dependant operons *litRQ* and *litAB* involved in carotenoid biosynthesis. Additionally, the expression levels of the ECF sigma factor, *litS* which is part of the *litAB* operon, remained unchanged. One possible explanation is the absence of RbpA or RbpA that is not capable of binding to HrdB dependant promoters in the S401 (pRT802) and S401 (pRT802::*rbpA*^{3KRA}) could increase the competition between the principle sigma factor, HrdB with other sigma factors like LitS. Thus resulting in more LitS binding with RNAP and activation of the *crt* biosynthetic gene cluster. To confirm this, western blot analysis would be required using the same basis as the SigE detection but using a specific LitS derived antibody.

5.5.4 Future work

To establish a growth phenotype of the M85 mutants, firstly growth curves using optical density are required. Next EMSA or BACTH assays could be used to establish binding between the RbpA M85 residue and the corresponding residue in σ . During the construction of the M85 mutants, we also aimed to explore the importance of the RbpA N-terminus. Scanning alanine mutagenesis had identified that the L6 residue in the N-terminus was well-conserved and thus an RbpA^{L6A} allele was created. This mutant did show a dominant phenotype, despite this the results were not conclusive. Therefore, as future directions, both the C and N-terminus require further exploration to understand the alternative function of RbpA.

To identify the transcriptional changes that occur in response to ACT production of S401 (pRT802::*rbpA*^{WT}) and S401 (pRT802) we will need to re-culture the strains for 72-150 h to allow production of ACT and then to extract the RNA to perform qRT-PCR for ACT associated

genes. Lastly, all of the targets discussed in this chapter will require validation using techniques such as qRT-PCR.

Results chapter IV: Investigating the
relationship between *S. coelicolor*
RbpA and CarD *in vivo*

6.1.0 Overview

Previous multi-round *in vitro* transcription assays tested transcription at three ribosomal RNA promoters, P2, P3 and P4. Promoter P2 was constitutively expressed, whilst no transcription was observed at the P3 and P4 in the absence of RbpA and/or CarD suggesting that the transcription factors might have overlapping function (L. Humphrey and M. Paget, personal communication). Indeed, this might explain why mutation of key residues that contact DNA in both proteins do not lead to null mutation phenotypes; for example, a CarD W86A mutation, which should be affected in DNA binding, might be masked by the presence of RbpA that also stabilises open complexes. Consistent with this, a ChIP-Seq experiment revealed co-localisation of RbpA and CarD at all σ^{HrdB} -dependent promoters (Laurence Humphrey and M. Paget, personal communication) and a recent structural model has suggested that RbpA and CarD are able to simultaneously contact DNA upstream of the -10 promoter region from opposite directions (Hubin et al., 2017a). Therefore, this chapter set out to identify the possible overlapping role between the two RNA polymerase binding proteins *in vivo*, by developing a strain in which both proteins are co-depleted.

6.1.1.0 Generation of stable RbpA-DAS and CarD-DAS strains

RbpA and CarD are both essential proteins and therefore it is not possible to make stable deletions. Conditional mutants were generated for both transcriptional regulators, by placing wild-type alleles under the control of a thiostrepton-inducible promoter, while deleting the gene at the native locus. In theory, depletion of each protein could be enacted by removal of the inducer. However, potential differences in stability would result in differences in the rate of depletion which would complicate the experiment. It was therefore decided to co-ordinately deplete the proteins using the same method of directed proteolysis, by tagging each gene at the native chromosomal locus with a proteolytic “DAS” tag (Figure 6.2). This system of proteolytic degradation has already been used in several organisms such as *Bacillus* (Griffith & Grossman, 2008), *E. coli* (McGinness et al., 2006), *M. tuberculosis* (Kim et al., 2011) and has been modified for *S. coelicolor* (L. Humphrey and M. Paget, personal communication). The “DAS+4” proteolysis system has been developed from a natural proteolytic mechanism. There are several proteolysis pathways, each playing key roles in natural protein turnover. A

widespread system is the ClpXP protease system, which is involved in the degradation of misfolded proteins and the regulation of proteins responding to stresses (Gottesman et al., 1998). ClpX is a hexameric AAA+ ATPase which recognises and unfolds proteins for degradation (Wawrzynow et al., 1995), delivering them for degradation by ClpP, a serine peptidase (Flynn et al., 2003). Collectively, different AAA+ ATPases recognise short peptide sequences so the specificity of the degradation event can be increased by using an N-terminal or C-terminal tag; C-terminal tags are favourable as they reduce interference of the protein function (Sauer et al., 2004). One such C-terminal tag is the 11-residue peptide *ssrA*-tag (AANDENYALAA), which is co-translationally added to peptide chains by a transfer-messenger RNA (tmRNA), when ribosomes stall at, for example, truncated 3' mRNA ends (Keiler et al., 1996). An adaptor protein SspB, helps to deliver the substrate to ClpXP by binding to the N-terminal sequence of the SsrA tag, while ClpX recognises the three C-terminus residues (Flynn et al., 2001). In theory, regulation of *sspB* could therefore determine the rate of degradation - however, in *E. coli*, ClpXP can degrade an SsrA-tagged protein in the absence of SspB (Kim et al., 2011; McGinness et al., 2006). To address this, and develop a controllable system, the SsrA-tag was changed to AANDENYSENYADAS (named DAS+4), which renders the tagged protein completely dependent on SspB for degradation (McGinness et al., 2006). SspB does not occur in Actinobacteria but has been shown to work in *Streptomyces*, where it was placed under the control of the thiostrepton-inducible *tipA* promoter (L. Humphrey and M. Paget, personal communication).

Two stage PCR was implemented to introduce the proteolytic DAS+4 tag, AANDENYSENYADAS, at the C-terminus of *rbpA* and *carD*, immediately upstream from the stop codons of each. The first stage amplifies two 1,180 bp flanks, the oligonucleotides that provide coverage over the DAS+4 tag overlap by 22 bases (Table 2.11). The second stage uses the two PCR products from the first stage and two new internal oligonucleotides (one forward for the first PCR product and the reverse for the second PCR product) to amplify a *rbpA*-DAS+4 PCR product of 1,960 bp and *carD*-DAS+4 of 1,955 bp (Table 2.11). The PCR products (*rbpA*-DAS+4 and *carD*-DAS+4) with 500 bp upstream and downstream, was ligated into pBluescript II SK+ and were confirmed by sequencing.

Traditionally, two-stage homologous recombination using non-replicating plasmids is used to introduce stable mutant alleles into the genome (Kieser et al., 2000). The first cross-over event involves integrating the mutant allele via one of the large flanking regions upstream or

downstream of the mutation, with a vector-borne antibiotic marker providing selection. The second cross-over event is identified by the loss of the resistance marker, the frequency of which is very low (<1%), making this a time-consuming screening method. Therefore, Fernandez-Martinez and Bibb (2014) developed the Meganuclease system for Actinobacteria to ease the isolation of stable, marker-less mutants and was slightly modified from Lu et al (2010). In this system, the non-replicating plasmid (pIJ2738) contains an 18 bp recognition site (5'-TAGGGATAACAGGGTAAT) for the *Saccharomyces cerevisiae* meganuclease I-SceI, which is a 235-amino acid monomeric endonuclease involved in mating type switching. Following the first cross-over event, a plasmid (pIJ2742) that expresses the meganuclease, I-SceI is introduced, which leads to lethal double strand breaks in the chromosome of any strain containing the 1st cross-over. The *rbpA*-DAS+4 and *carD*-DAS+4 alleles were isolated as NotI/SpeI fragments and cloned into pIJ2738 that had been digested with the same enzymes to give pIJ2738::*rbpA*-DAS+4 and pIJ2738::*carD*-DAS+4. The recombinant plasmids were used to transform *E. coli* ET12567/ pR9406, and then conjugated into *S. coelicolor* S101 (Δ *rbpA*), J1915 (M145 Δ *glkA*, the parent of S101) and M145 (wild type; Fig. 6.1). S101 grows slowly and forms small colonies compared to M145, however first cross-over recombinants reverted to a wild-type phenotype, which confirmed that the *rbpA*-DAS+4 allele was functional for RbpA activity. After making spore stocks for these strains, pIJ2742, which expresses the I-SceI meganuclease, was conjugated into the first cross-over strains using thiostrepton for selection, and induction of the nuclease. The ex-conjugants were sub-cultured to single colonies on MS agar containing 15 μ g/mL thiostrepton for the purification of the second cross-over recombinants, and then screened using colony PCR for the presence of the DAS+4 tag. The colony PCR primers were designed to amplify the open reading frame of *rbpA* and *carD* and the DAS+4 tag, recombinants were confirmed upon a small shift of 45 bp compared to *rbpA* and *carD* without the DAS+4 tag (Table 2.11). Upon confirmation, these colonies were plated onto MS agar in the absence of selection and incubated at 37°C for 4-5 days, to ensure the loss of pIJ2742, which contains the temperature-sensitive replication origin of pSG5. The potential marker-less stable mutants were checked for the absence of resistance genes by streaking to MS agar supplemented with apramycin or thiostrepton, at 30°C for 4-5 days, followed by production of spore stocks. The resulting strains were named S407 (M145 *rbpA*^{DAS+4}); S408 (J1915 *rbpA*^{DAS+4}); S409 (M145 *carD*^{DAS+4}); S410 (J1915 *carD*^{DAS+4}); S411 (J1915 *rbpA*^{DAS+4}/*carD*^{DAS+4}). As was the case with *rbpA*, *carD*^{DAS+4} strain growth on solid media was

indistinguishable from that of the parent strain, suggesting that the tag had little if any impact on protein function. In order to target the DAS-tagged proteins for degradation, pIJ6902::*sspB*, whereby *sspB* expression is controlled by the thiostrepton-inducible *tipAp* promoter, was conjugated into each of the strains. The organisation of the engineered strains is summarised in Figure 6.2.

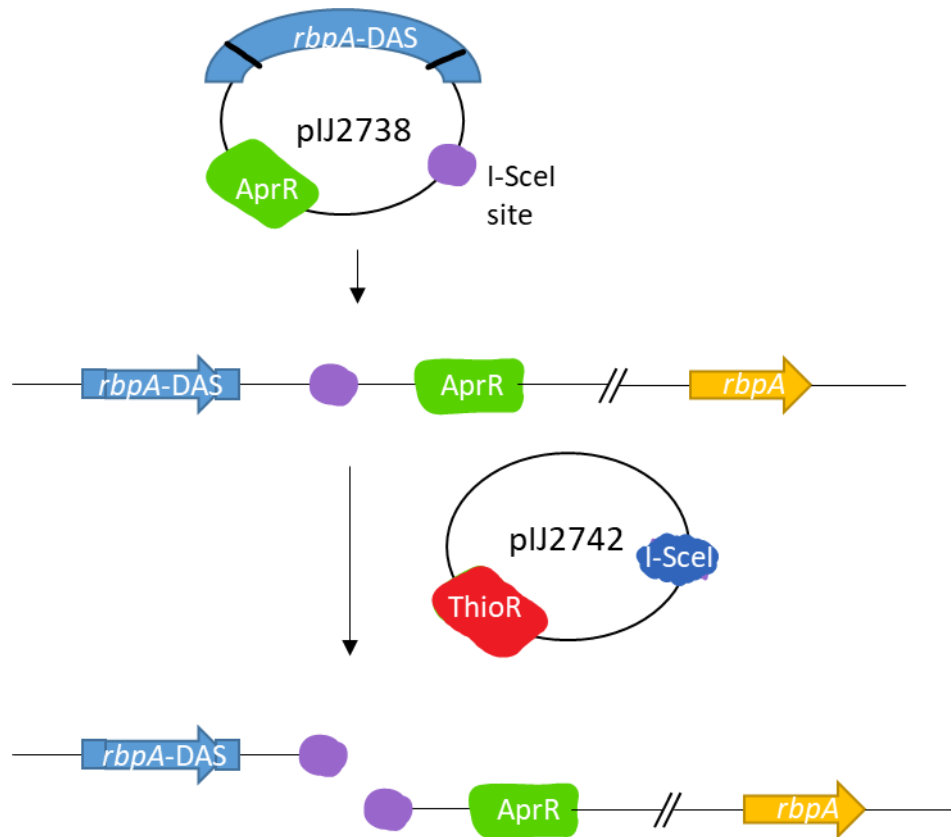


Figure 6.1. The Meganuclease system to generate marker-less gene replacements in *S. coelicolor*. pIJ2738, which carries apramycin resistance, the Sce-I recognition site and the *rbpA*^{DAS+4} or *carD*^{DAS+4} alleles flanked by 500 bp DNA, was conjugated into *S. coelicolor* J1915, M145 or S101. Single exconjugants were selected for conjugation with pIJ2742 which expresses the I-SceI meganuclease, controlled thiostrepton-inducible *tipAp* promoter. An overlay 15 µg/mL thiostrepton was applied to the plates to select for exconjugants that had undergone the second cross-over event, and therefore lost the Sce-I recognition site. Colonies confirmed by PCR for the presence of the DAS tag were cultured at 37°C, which causes the loss of the temperature-sensitive pIJ2742, and this was confirmed by testing for absence of growth on thiostrepton. This figure shows an example of the meganuclease system using *rbpA*^{DAS}, the

same was carried out for *carD*^{DAS}. This protocol has been modified from Fernandez-Martinez and Bibb (2014).

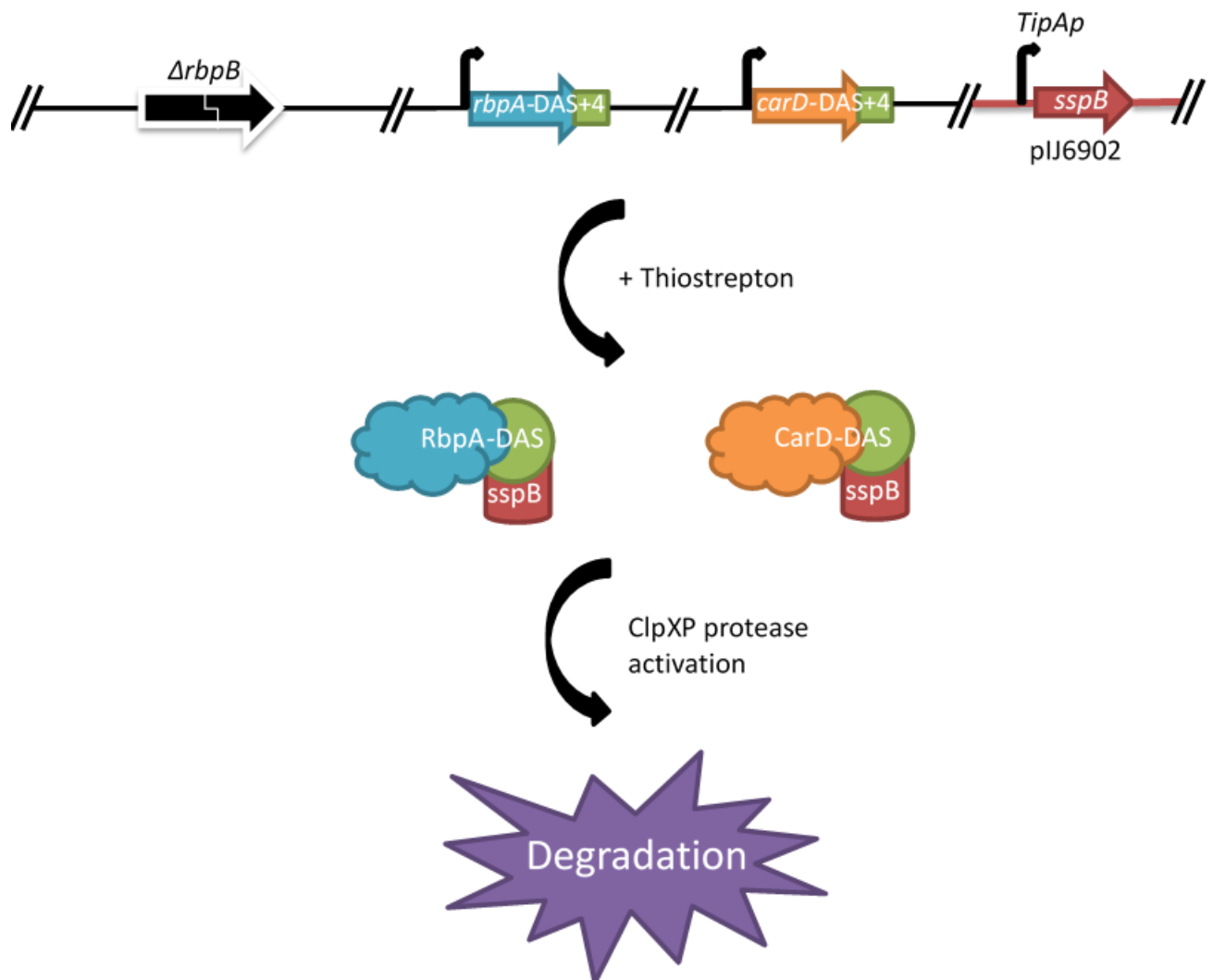


Figure 6.2. Summary of the DAS+4 depletion system. The example given is S403 in which both *rbpA-DAS+4* and *carD-DAS+4* alleles are present, each at the native locus. The strain also contains a stable *rbpB* deletion and *pIJ6902::sspB* integrated at the ϕ C31 attachment site. The addition of thiostrepton activates *tipAp* causing expression of *sspB*. The adapter protein recognises the C-terminal DAS+4 tags in RbpA and CarD, leading to degradation via ClpXP protease.

6.1.2.0 Validation of sspB-dependent depletion

The DAS dependant depletion of RbpA and CarD on a protein level was first validated by inserting a FLAG tag at the 3' end of the RbpA and CarD ORFs, but upstream of the DAS tag, using western blot analysis.

6.1.2.1 Construction of the FLAG-DAS+4- tag strains

To mimic the previously constructed DAS-tagged strains, an in-frame 3XFLAG tag was introduced immediately upstream of the C-terminal DAS+4 tags in *rbpA* and *carD*. Inverse PCR mutagenesis was performed using pIJ2738::*rbpA*-DAS+4 and pIJ2738::*carD*-DAS+4 as templates and oligonucleotides that introduced the 3xFLAG sequence, DIDYKDDDDKGS, into *rbpA* (R/C_FLAGDASinvFand RbpAFLAGDASinvR; Table 2.12) and *carD* (R/C_FLAGDASincF and CarDFLAGDASinvR; Table 2.12). Following sequence confirmation, *rbpA*-FLAG-DAS+4 and *carD*-FLAG-DAS+4 were cloned as SpeI/NotI fragments into pRT802, which was transformed into *E. coli* ETR and conjugated into *S. coelicolor* J1915 (pIJ6902) and J1915 (pIJ6902::*sspB*). To test for SspB-dependent proteolysis, spores were cultured to OD₄₅₀ 0.8, a time point 0 min sample taken, and then thiostrepton (15 µg/mL, final concentration) was added to the cultures, before samples for soluble extract preparations were taken over a time course. Initially time points 0, 3, 6 and 22 h were used, but no protein was detected after 0 h. A shorter time-course of 0, 15, 30, 60 and 120 min was repeated and revealed that protein depletion was only detected at the last time point. Therefore, samples were taken at 0, 1, 2, 4 and 6 h after thiostrepton treatment and RbpA^{FLG-DAS+4} and CarD^{FLG-DAS+4} levels were quantified at each time point (Fig. 6.3 and 6.4). Both engineered constructs demonstrate clear depletion of DAS-tagged RbpA and CarD upon addition of thiostrepton, whereas no depletion was observed in the absence of thiostrepton, implying that degradation is mediated by SspB. It appears however that depletion takes upwards of 4-6 h, which has important implications for studying the effects of RbpA/CarD depletion on the transcriptome. In the J1915 (pIJ6902::*sspB*, pRT802::*carD*-FLAG-DAS) strain low levels of CarD were detected even at time 0 min compared to the control, although this appeared to decrease upon thiostrepton addition. These data suggest that relatively weak effect of SspB-mediated degradation of CarD on growth does account for poor depletion.

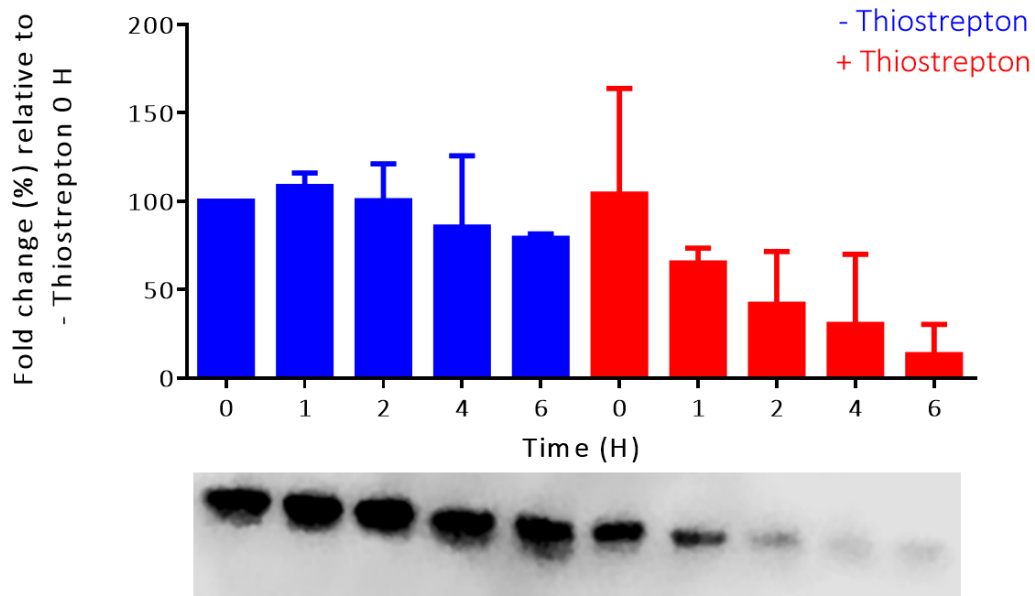


Figure 6.3. *sspB*-mediated depletion of RbpA in *S. coelicolor* J1915 (pIJ6902::*sspB*, pRT802::*rbpA*-FLAG-DAS+4). Two cultures of the strain were grown to OD₄₅₀ 0.8, and 15 µg/mL thioestrepton was added to one flask but not the other. 5 mL samples at 0, 1, 2, 4, and 6 h were harvested by centrifugation and soluble cell lysates prepared by sonication and centrifugation (Methods, 2.5.3). Extracts were normalised using a Qubit HS protein assay before separation on a 10% SDS-PAGE gel, followed by western blotting. The protein was detected using an anti-FLAG antibody (Sigma). The data are compiled from three biological replicates and represents fold-change with respect to levels determined in the absence of thioestrepton at time point 0 min.

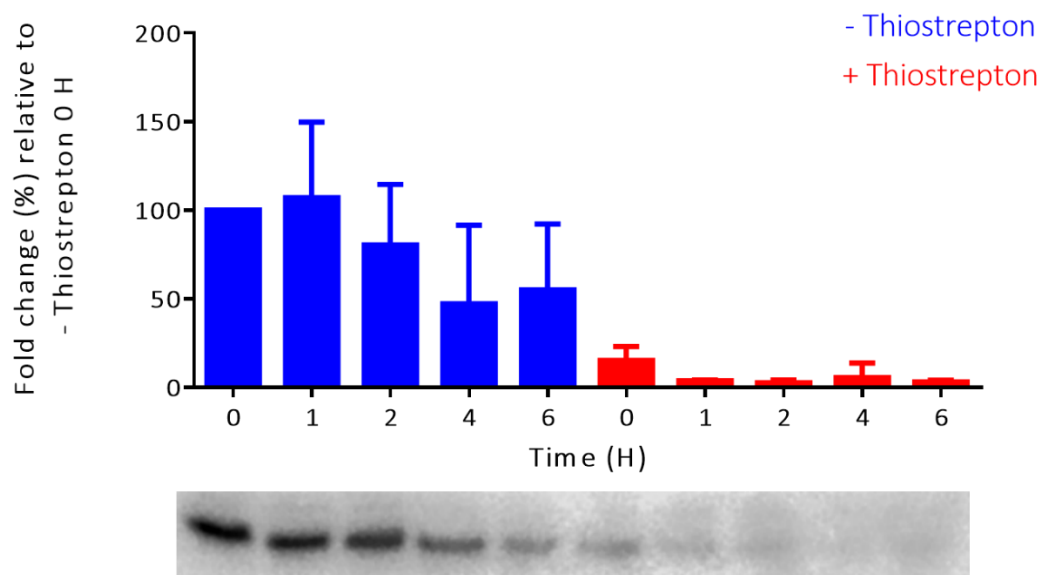


Figure 6.4. *sspB*-mediated depletion of CarD in *S. coelicolor* J1915 (pIJ6902::*sspB*, pRT802::*carD*-FLAG-DAS+4). J1915 (pIJ6902::*sspB*, pRT802::*carD*-FLAG-DAS+4) was cultured to OD450 0.8 in the presence/absence of 15 µg/mL thiostrepton. 5 mL samples at the time points 0, 1, 2, 4 and 6 h harvested by centrifugation and the cell lysate was normalised using the Qubit HS protein assay before running on a 10% SDS-PAGE, followed by western blotting. This data is compiled from three biological replicates and the data represents fold change with respect to J1915 (pIJ6902::*sspB*, pRT802::*carD*-FLAG-DAS+4) in the absence of thiostrepton at time point 0 min.

6.1.1.1 Depletion of RbpA-DAS+4 and CarD-DAS+4 in a *sspB*-dependant manner

Initial studies were carried out with DAS+4-tagged strains constructed in the J1915 background in which *rbpB* is present. Strains S408 (*rbpA*^{DAS+4}), S410 (*carD*^{DAS+4}), S411 (*rbpA*^{DAS+4} / *carD*^{DAS+4}) and J1915, each additionally contained pIJ6902::*sspB*, and were cultured on MS agar containing 0-15 µg/mL thiostrepton at 30°C for 3-4 days to investigate dose responsiveness of *sspB*-dependant depletion (Fig. 6.5). In the case of S408 and S411(pIJ6902::*sspB*), thiostrepton inhibited growth significantly compared to J1915 (pIJ6902::*sspB*) suggesting that RbpA^{DAS+4} was effectively depleted. However, for S410 (pIJ6902::*sspB*), inhibition of growth was only slightly greater than that of the control, suggesting either that CarD^{DAS+4} was not depleted, or that the CarD depletion did not have significant effect on growth. As expected, the growth of S411 (pIJ6902::*sspB*), in which both *rbpA* and *carD* are DAS-tagged was inhibited by thiostrepton,

but no more so than S408 in which only RbpA is tagged. Therefore, there did not appear to be a cumulative effect of RbpA^{DAS+4} and CarD^{DAS+4} co-depletion on growth. However, the data also raised the possibility that CarD^{DAS+4} depletion was not as effective as that of RbpA^{DAS+4}.

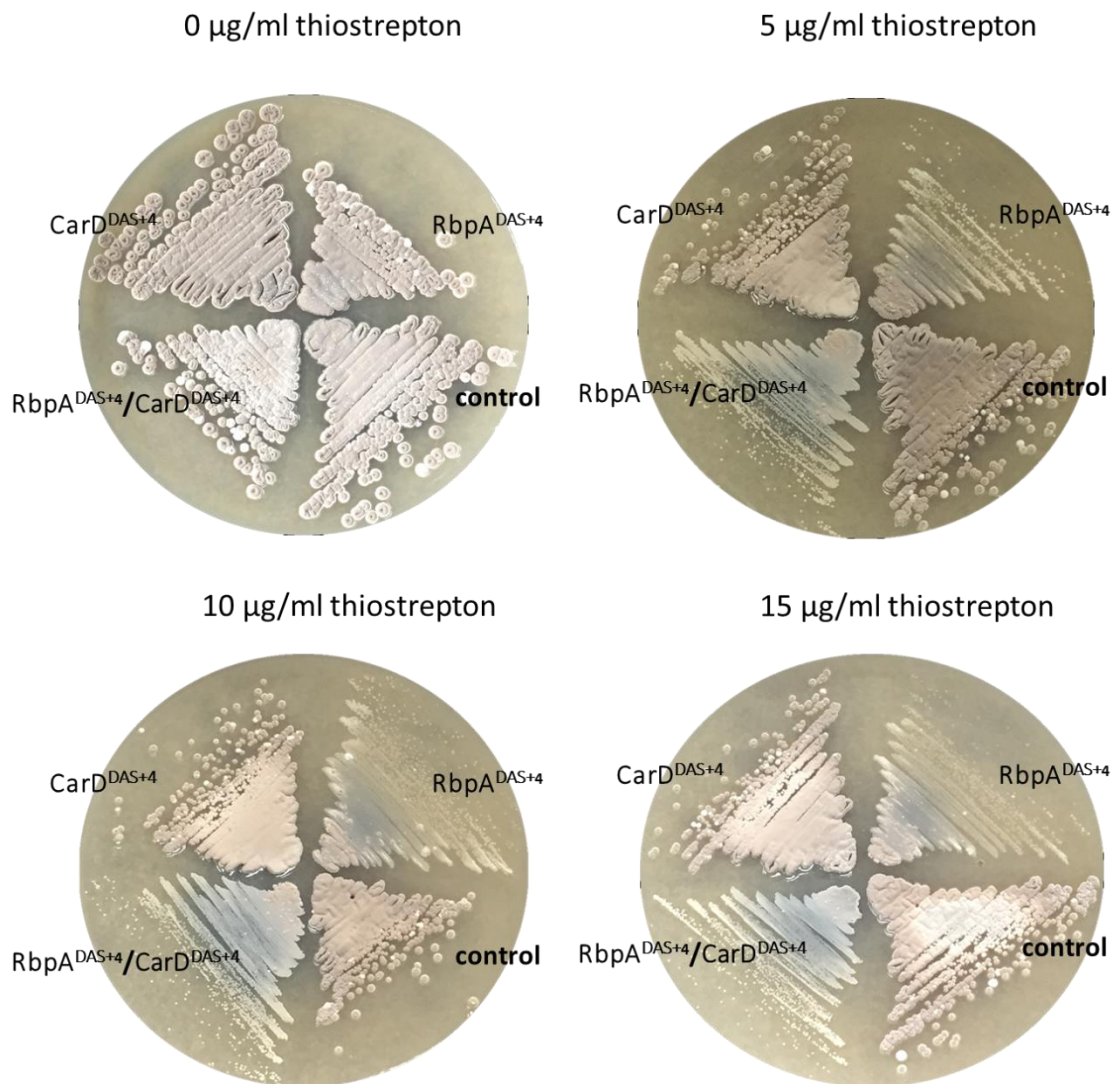


Figure 6.5. The effects of *sspB*-dependent depletion of *rbpA* and *carD* on growth. The dose *sspB* dependent depletion of S408 (*rbpA*-DAS+4), S410 (*carD*-DAS+4) and S411 (*rbpA*-DAS+4/*carD*-DAS+4) with each additionally containing pIJ6902::*sspB*. *sspB* is induced using 0-15 µg/mL thiostrepton which activates the *tipA* promoter *sspB*. The control does not contain any DAS+4 tagged proteins.

6.1.2.2 The growth phenotype of the DAS+4 depletion strains are further amplified with $\Delta rbpB$

SspB-mediated depletion of RbpA did not completely inhibit growth, which is probably explained by the presence of *rbpB*. As described earlier, *rbpB* has an overlapping function with *rbpA* and together they confer an essential role in *S. coelicolor*. Therefore, to completely deplete RbpA/RbpB activity, DAS-tagged strains were re-constructed in a $\Delta rbpB$ derivative of *S. coelicolor* J1915, S406 (A. Tabib-Salazar and M. Paget, personal communication). The gene replacement plasmids pIJ2738::*rbpA*-DAS+4 and pIJ2738::*carD*-DAS+4 were introduced into S406 and the second cross-over recombination events again forced using pIJ2742, as described above and the new strains designated S414 (J1915 $\Delta rbpB$ *rbpA*-DAS+4, pIJ6902::*sspB*), S415 (J1915 $\Delta rbpB$ *carD*-DAS+4, pIJ6902::*sspB*) and S416 (J1915 $\Delta rbpB$ *rbpA*-DAS+4/*carD*-DAS+4, pIJ6902::*sspB*) were introduced with pIJ690::*sspB*. The $\Delta rbpB$ DAS strains showed reduced growth compared to the presence of a wild-type *rbpB* (Fig 6.6A and 6.6B). The growth was almost completely redundant in the S414 ($\Delta rbpB$ *rbpA*-DAS+4, pIJ6902::*sspB*) and S416 ($\Delta rbpB$ *rbpA*-DAS+4/*carD*-DAS+4, pIJ6902::*sspB*) strains whilst a small difference in growth was shown for S415 ($\Delta rbpB$ *carD*-DAS+4, pIJ6902::*sspB*) strain when comparing to the respective wild-type *rbpB* strains. The control strain J1915 (pIJ6902::*sspB*) demonstrated a small reduction in colony size and growth which suggested that *sspB* or the inducer thiostrepton have non-specific action to some extent. These same strains in the $\Delta rbpB$ genetic background were plated onto MS agar containing 0 $\mu\text{g/mL}$, 5 $\mu\text{g/mL}$, 10 $\mu\text{g/mL}$ and 15 $\mu\text{g/mL}$ (Fig. 6.7). The biggest difference in growth exists between 0 $\mu\text{g/mL}$ and 15 $\mu\text{g/mL}$. The depletion of RbpA^{DAS+4} alone produces ACT whilst the co-depletion of RbpA^{DAS+4}/CarD^{DAS+4} produces noticeably more ACT but the depletion of CarD^{DAS+4} is slow and has little effect between 5-15 $\mu\text{g/mL}$. These data eluded to RbpA having more of a dominant effect on growth compared to CarD.

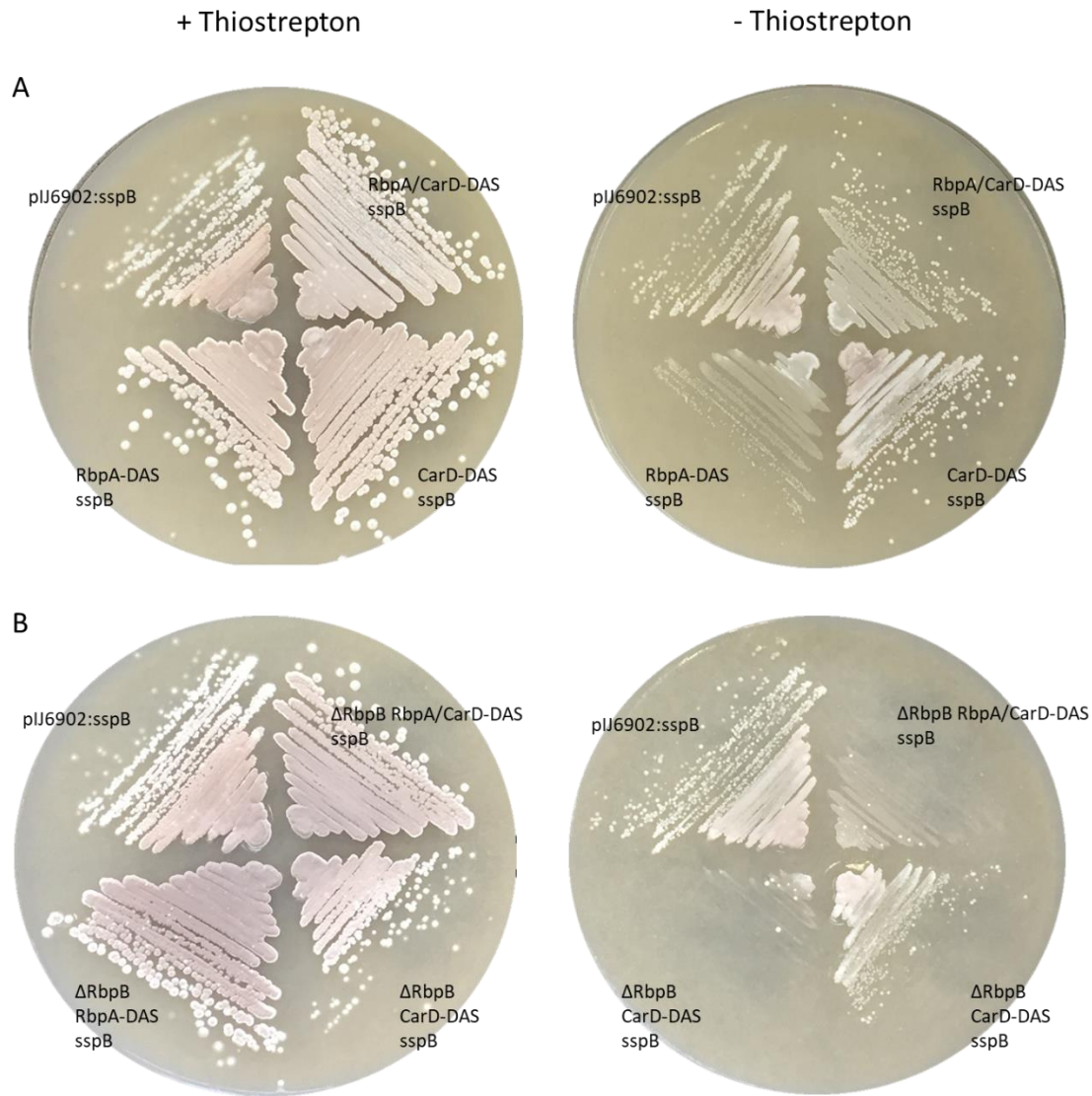


Figure 6.6. The effect of deletion of *rbpB* from the DAS+4 tagged *rbpA* and *carD*. (A) left: without thiostrepton induction of *sspB*, plated J1915 (pIJ6902::sspB) control, S411 (*rbpA*-DAS+4/*carD*-DAS+4, pIJ6902::sspB), S412 (*rbpA*-DAS+4, pIJ6902::sspB) and S413 (*carD*-DAS+4, pIJ6902::sspB). Right: the same strains as the left plate, with 20 μ g/mL thiostrepton. (B) left: without thiostrepton induction of *sspB*, plated J1915 (pIJ6902::sspB), S416 (Δ *rbpB* *rbpA*-DAS+4/*carD*-DAS+4, pIJ6902::sspB), S406 (Δ *rbpB* *rbpA*-DAS+4, pIJ6902::sspB) and S406 (Δ *rbpB* *carD*-DAS+4, pIJ6902::sspB). Right: the same strains as the left plate, with 20 μ g/mL thiostrepton. These plates were cultured at 30°C for 72 h.

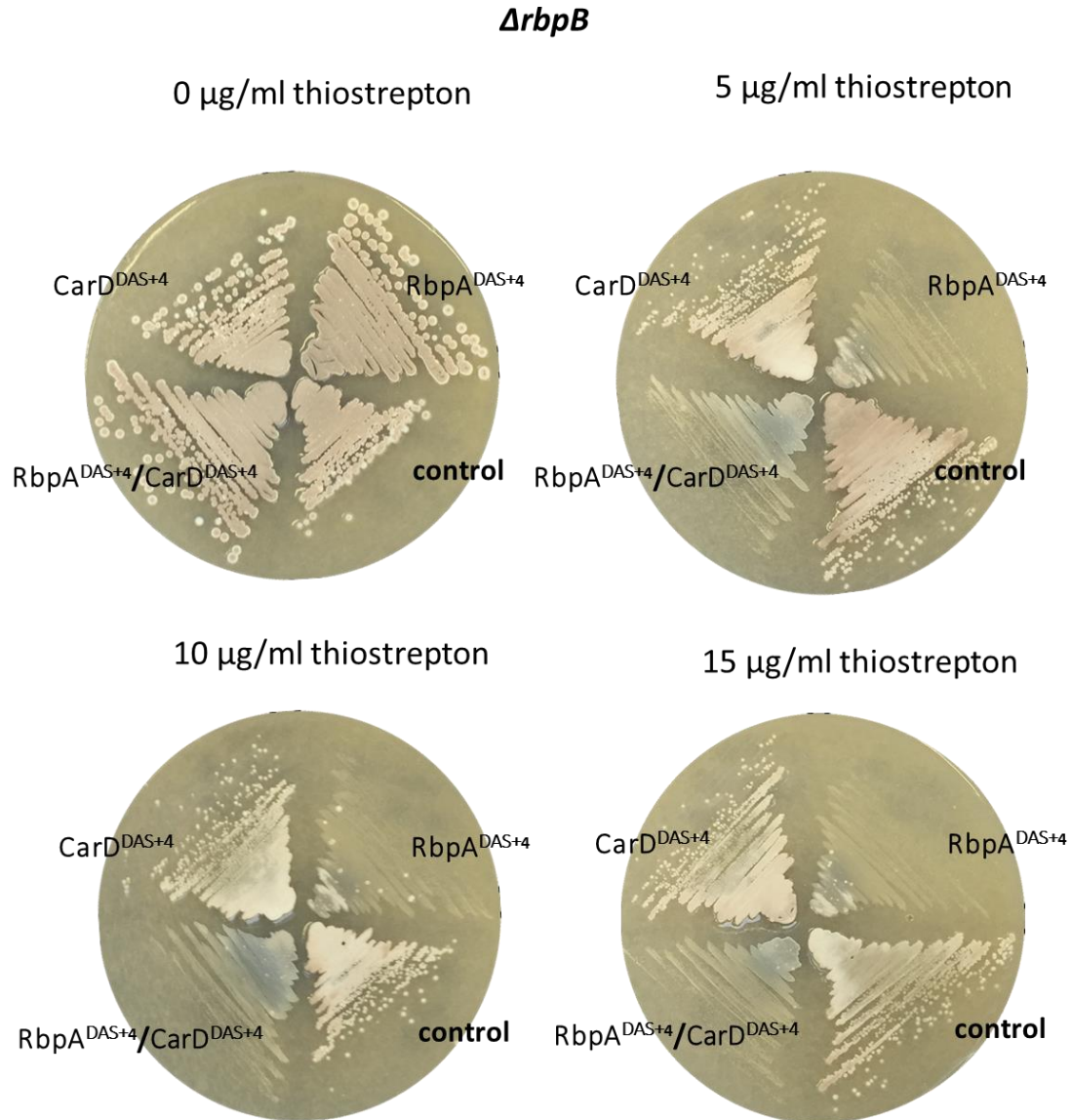


Figure 6.7. The effects of *sspB*-dependent depletion of *rbpA* and *carD* on growth in a $\Delta rbpB$ null background. The dose *sspB*-dependent depletion of S414 $\Delta rbpB$ (*rbpA*-DAS+4), S415 $\Delta rbpB$ (*carD*-DAS+4) and S416 $\Delta rbpB$ (*rbpA*-DAS+4/*carD*-DAS+4) with each additionally containing pIJ6902::*sspB*. *sspB* is induced using 0-15 $\mu\text{g/ml}$ thiostrepton which activates the *tipA* promoter *sspB*. The control does not contain any DAS+4 tagged proteins.

6.1.2.3 Depletion of the DAS tagged proteins in liquid culture

The initial studies of the *sspB*-dependent depletion of RbpA and CarD on solid MS agar confirmed a growth phenotype. This growth phenotype was further investigated by generating growth curves from liquid culture of the J1915 $\Delta rbpB$ (pIJ6902::*sspB*) DAS+4 strains. Each strain

was cultured in 2 flasks, 15 µg/mL thiostrepton was added to flask one at OD₄₅₀ 0.8 whilst no thiostrepton was added to flask two and this was repeated to obtain results for 3 biological replicates. Samples were taken at regular intervals up to 64 h and the optical density was measured (Fig. 6.8). The S414 ($\Delta rbpB$ *rbpA*-DAS+4, pIJ6902::*sspB*) and S416 ($\Delta rbpB$ *rbpA*-DAS+4/*carD*-DAS+4, pIJ6902::*sspB*) strains demonstrated a clear difference between the -/+ thiostrepton conditions. This confirmed that *sspB* is activated at the *tipA* promoter in the presence of thiostrepton, both RbpA and CarD are essential for growth, depletion of both reduces the growth rate of *S. coelicolor*. The difference is not completely clear for the S415 ($\Delta rbpB$ *carD*-DAS+4, pIJ6902::*sspB*). The results of the liquid culture growth curves do confirm the western blot analysis of the *sspB* time dependant depletion of the DAS+4 tagged proteins. However, the error bars are large, it has been difficult to replicate the same conditions for each flask, this is because of *Streptomyces*' ability to produce mycelium which clump differentially and attempts to reduce the effect of clumping by placing a sterile spring into the flasks and siliconizing the flasks with Sigmacote (Sigma-Aldrich) were made.

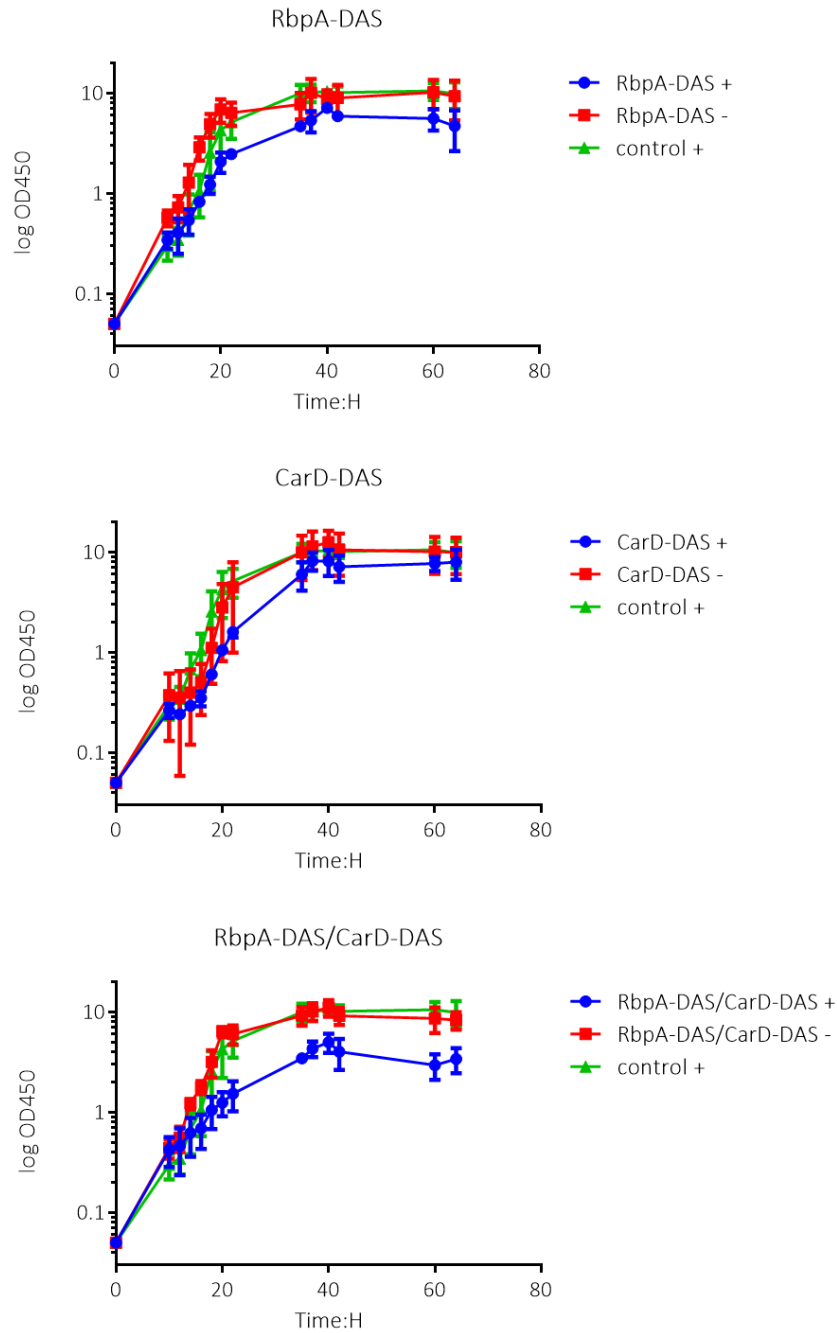


Figure 6.8. *sspB*-dependant depletion of **RbpA** and **CarD** in liquid media. Spore stocks for the S414 ($\Delta rbpB$ *rbpA*-DAS+4, pIJ6902::*sspB*; top), S415 ($\Delta rbpB$ *carD*-DAS+4, pIJ6902::*sspB*; middle), S416 ($\Delta rbpB$ *rbpA*-DAS+4/*carD*-DAS+4, pIJ6902::*sspB*; bottom) strains were normalised and pre-germinated for 3 h before culturing for up to 64 h in YEME supplemented with glycerol. For each strain, *sspB* was induced with 15 $\mu\text{g}/\text{mL}$ thiostrepton in the first flask (blue) at OD450 0.8 whilst the second flask contains no thiostrepton (red). J1915 (pIJ6902::*sspB*) serves as a control (green). The data is compiled from three biological replicates.

6.2.0 Testing for the relationship between *rbpA* and *carD*

The CarD^{W86A} mutant has been extensively researched in *Mycobacteria*, however relatively little is known about the CarD^{W85A} function in *S. coelicolor*. The W85A residue lies within the CarD C-terminal domain and appears close to the DNA minor groove at around -12 bp promoter region (Srivastava et al., 2013). *In vitro* studies demonstrated that the alanine substitution of W85, was unable to activate transcription (Srivastava et al., 2013). Contrastingly, CarD^{W85A} was shown to bind to DNA, and associate with RNA polymerase however demonstrated a reduced growth rate in liquid media and reduced pigmentation on solid media in *M. smegmatis* (Garner et al., 2014). The true function of the tryptophan residue in CarD therefore remains unclear. The *rbpA* BL, DNA-binding determinant has been extensively described in results chapter I. RbpA and CarD have very well conserved DNA-binding regions which may be important for binding at the same DNA promoters, therefore the co-depletion strain S416 ($\Delta rbpB$ *rbpA*-DAS+4/*carD*-DAS+4 (pIJ6902::*sspB*) was implemented in this study to investigate the growth phenotype of various combinations of the RbpA and CarD DNA-binding mutant alleles.

6.2.1 Construction of the RbpA BL and CarD^{W85A} integrative plasmids

The DAS+4 strains already have integration of the pIJ6902::*sspB* at the phiC31 site, therefore mutant alleles were cloned into pRT802 which integrates at the phiBT1 *attP* site. Gibson assembly was used to insert *rbpA* with 100 bp upstream and downstream DNA using primers, 1-pRT802:*rbpA* F and 2-*carD*-*rbpA* R (Table 2.13) and *carD* with 150 bp upstream and 100 bp downstream DNA into *SpeI* digested pRT802 using primers 3-*rbpA*-*carD* F and 4-*carD*-pRT802 R (Table 2.13).

The primers designed for this amplification, were used to engineer the following plasmids:

- 1) pRT802::*rbpA*^{WT}*carD*^{WT}
- 2) pRT802::*rbpA*^{R80A}*carD*^{WT}
- 3) pRT802::*rbpA*^{3KA}*carD*^{WT}
- 4) pRT802::*rbpA*^{WT}*carD*^{W85A}
- 5) pRT802::*rbpA*^{R80A}*carD*^{W85A}
- 6) pRT802::*rbpA*^{3KA}*carD*^{W85A}

7) pRT802

The constructs were confirmed with sequencing and transformed into *E. coli* ETR and conjugated into S416 ($\Delta rbpB$ *rbpA*-DAS+4/*carD*-DAS+4 (pIJ6902::*sspB*), Figure 6.10 summarizes the engineered elements of this strain. To test the phenotypes of the engineered constructs, spores were plated in the absence/presence of 15 $\mu\text{g}/\text{mL}$ thiostrepton on MS agar (Fig. 6.11).

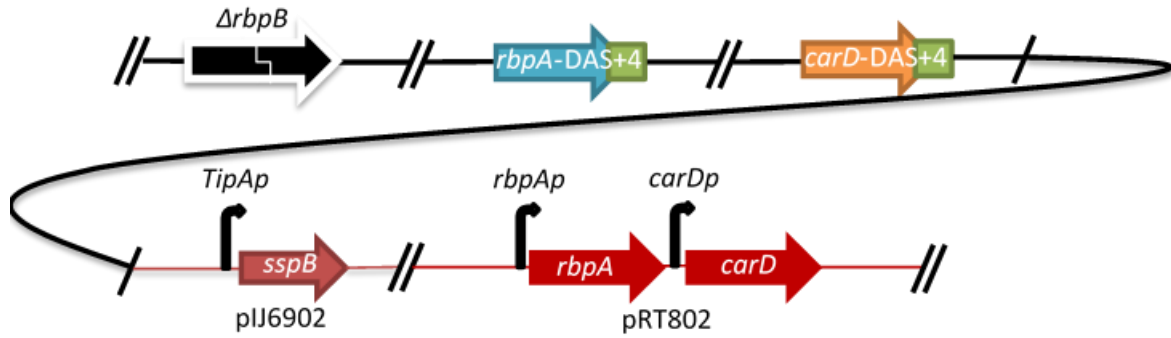


Figure 6.10. The engineered elements of the *rbpA*-DAS+4 and *carD*-DAS+4 strains with pRT802. Constructs are engineered with $\Delta rbpB$, *rbpA*-DAS+4/*carD*-DAS+4 with integration of pIJ6902::*sspB* (orange) at ϕIC31 site, *sspB* is under the control of a *tipA* promoter induced by thiostrepton. The Gibson assembled *rbpA* and *carD* mutant alleles with native promoters are ligated to pRT802 (red) and are integrated at the $\phi\text{IBT1 attP}$ site. The pRT802 constructs are: pRT802::*rbpA*^{WT}*carD*^{WT}, pRT802::*rbpA*^{R80A}*carD*^{WT}, pRT802::*rbpA*^{3KA} *carD*^{WT}, pRT802::*rbpA*^{WT} *carD*^{W85A}, pRT802::*rbpA*^{R80A} *carD*^{W85A}, pRT802::*rbpA*^{3KA} *carD*^{W85A} and pRT802.

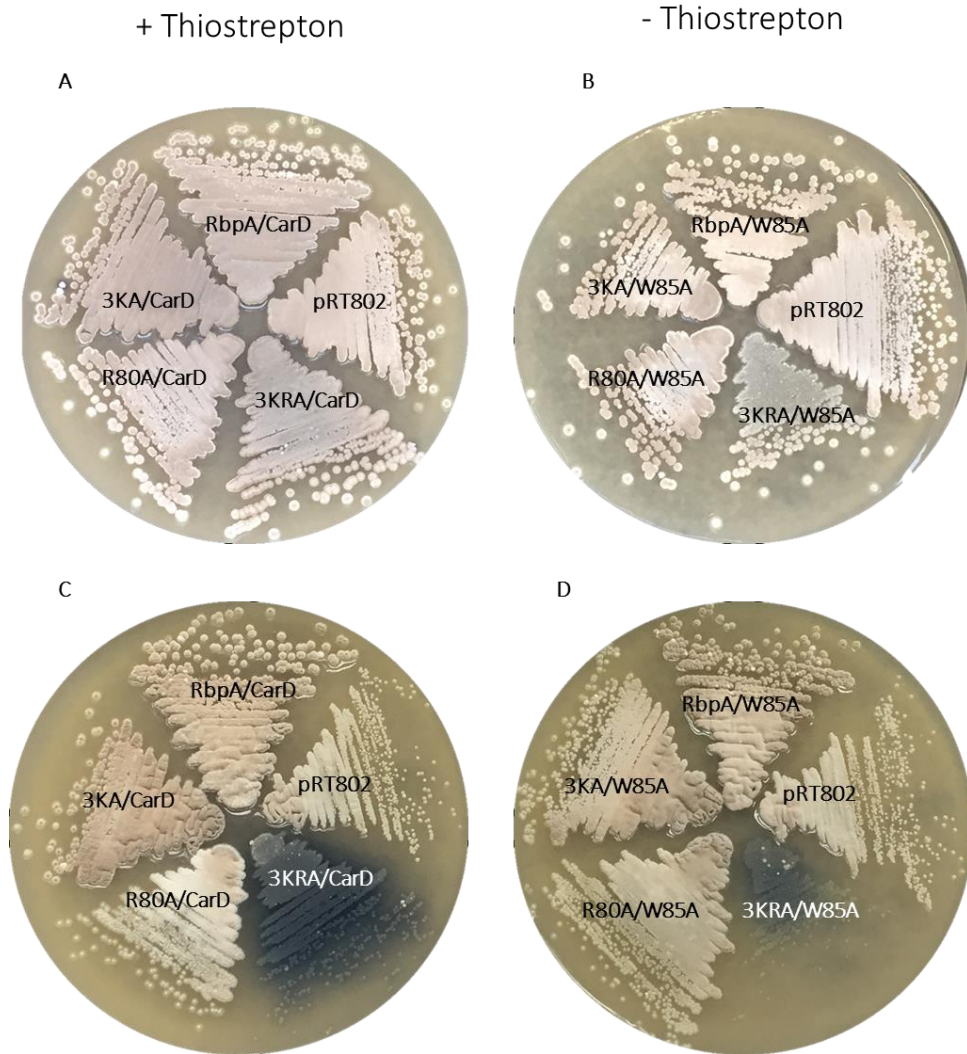


Figure 6.11. Phenotype of RbpA BL and CarD^{W85A} mutant alleles in the *sspB*-dependent co-depletion of *rbpA/carD* strain. The engineered constructs have been conjugated into S416 $\Delta rbpB$ *rbpA*-DAS+4/*carD*-DAS+4 via the *attP* integration site. Spore stocks for each of these constructs are plated onto MS agar in the presence/absence of 15 μ g/mL thiostrepton. pRT802::*rbpA*^{WT}/*carD*^{WT}, pRT802, pRT802::*rbpA*^{3KRA}/*carD*^{WT}, pRT802::*rbpA*^{R80A}/*carD*^{WT}, and pRT802::*rbpA*^{3KA}/*carD*^{WT} in the (A) absence and (C) presence of 15 μ g/mL thiostrepton. pRT802::*rbpA*^{WT}/*carD*^{W85A}, pRT802, pRT802::*rbpA*^{3KRA}/*carD*^{W85A}, pRT802::*rbpA*^{R80A}/*carD*^{W85A}, pRT802::*rbpA*^{3KA}/*carD*^{W85A} in the (B) absence and (C) presence of 15 μ g/mL thiostrepton.

S416 pRT802::*rbpA*^{WT}/*carD*^{WT} and S416 pRT802::*rbpA*^{WT}/*carD*^{W85A} appear to grow similarly which means that the loss of tryptophan 85 in CarD alone does not affect the growth phenotype. The growth phenotype for S416 (pRT802::*rbpA*^{3KRA}/*carD*^{WT}) was significant,

compared to S416 (pRT802::*rbpA*^{WT}/*carD*^{W85A}), suggesting an important role for RbpA DNA binding region compared to the DNA binding tryptophan of CarD. The growth phenotype is amplified in S416 (pRT802::*rbpA*^{3KRA}/*carD*^{W85A}) compared to S416 (pRT802::*rbpA*^{WT}/*carD*^{W85A}), indicated that both DNA binding determinants are required for growth of *S. coelicolor*. In addition to this, S416 (pRT802::*rbpA*^{3KRA}/*carD*^{WT}) complements, whilst S416 (pRT802::*rbpA*^{3KRA}/*carD*^{W85A}) is unable to complement supporting the idea that both DNA binding regions are important for growth, and that both transcriptional regulators have alternative roles, aside from the already elucidated role in growth.

The phenotypes of the *rbpA* BL mutant alleles (Figure 6.11C) further confirms the phenotypes of the same mutant alleles in the conditional *rbpA* strain, S401 (Results chapter I). As seen previously, in results chapter I, the BL mutant (pRT802::*rbpA*^{3KA}/*carD*^{WT}) does not show a clear phenotype, nor does this change with the *CarD*^{W85A}.

6.3.0 Discussion

6.3.1 Induced degradation of RbpA^{DAS+4} and CarD^{DAS+4}

RbpA and CarD are well conserved across the Actinobacteria and appear to play individually essential roles in transcription initiation, despite each contacting the upstream edge of the -10 element to stabilise transcription intermediates (Hubin et al., 2015; Srivastava et al., 2013). However, the finding that the essential function of RbpA is unlikely to be in contacting promoter DNA at this location presents a scenario where the RbpA and CarD are redundant in contacting -10 upstream edge, but that one or both have essential additional functions that cannot be compensated by the other. The aim of this chapter was to investigate redundancy between RbpA and CarD in interacting with the upstream edge of the -10 element by depleting the wild type proteins whilst expressing DNA-binding alleles of RbpA and CarD in different combinations. To co-ordinately deplete wild-type versions of RbpA and CarD, both genes were DAS+4-tagged at their native loci. The efficiency of the DAS+4-tag system was investigated by monitoring levels of RbpA^{FLG-DAS+4} and CarD^{FLG-DAS+4} upon thiostrepton-induced *SspB* expression. Depletion of RbpA^{FLG-DAS+4} was shown to occur within 1 h of *sspB* induction whilst the depletion of CarD^{FLG-DAS+4} was not so clear. Other studies have shown depletion times upon *sspB* induction of 12 h using a GFP-DAS system in *M. smegmatis* (Kim et al., 2011), 24 h in a

Rho-DAS system in *M. tuberculosis* (Botella et al., 2017) and 15 min in *E. coli* (McGinness et al., 2006). The efficient generation of the different DAS+4-tagged strains was enabled by the use of the meganuclease counterselection system (Fernández-Martínez & Bibb, 2014). Depletion of RbpA^{DAS+4} in a *rbpB*^{WT} background only slightly reduced growth, consistent with partial redundancy of RbpA and RbpB. Therefore, further experiments were conducted in a Δ *rbpB* mutant and confirmed that RbpA is essential for growth in this background. However, depletion of CarD^{DAS+4} in both *rbpB*⁺ and Δ *rbpB* led to only minimal inhibition of growth, which was surprising because previous depletion using vector-borne *carD*^{DAS+4} in a Δ *carD* background significantly impacted growth suggesting that *carD* was essential (L. Humphrey, personal communication). Interpretation was also complicated by a slight inhibition of growth caused by *sspB* induction at the higher thiostrepton concentration in the parental strain J1915 (*rbpA*^{WT}, *carD*^{WT}) suggesting that *sspB* may signal other proteins for degradation non-specifically. It is not clear why *carD*^{DAS+4} depletion failed to prevent growth on growth solid media, although one possibility is that it is expressed at a higher level from its native locus, preventing SspB from fully depleting the protein. Another possibility for incomplete CarD^{FLG-DAS+4} depletion could be a result of the absence of a terminator in the integrative plasmid, resulting in unstable mRNA (Nouaille et al., 2017). Additionally, to this, 420 bp upstream of the CarD ORF was used for cloning, in this study, one possibility is that CarD may contain some regulatory elements further upstream of the 420 bp chosen resulting in an unclear depletion. Growth in liquid media reflected that on solid, with the simultaneous depletion of RbpA^{DAS+4} and CarD^{DAS+4} inhibiting growth. Interestingly, however, depletion of only RbpA^{DAS+4} had less impact on growth suggesting some redundancy between RbpA and CarD.

6.3.2 Validation of the DAS system using integrative RbpA BL mutants

Thus far, this study has shown the successful implementation of the *sspB* dependant DAS depletion system in RbpA and CarD. The DAS depletion system was validated by integrating mutants important for DNA binding in RbpA and CarD. The residues important for RbpA DNA binding were chosen because these had been created using an independent conditional system, therefore comparing the phenotypes of the same mutants in two different systems would help validation. The depletion of RbpA and CarD with integrated BL/CarD^{WT} mutants

demonstrate the exact same phenotype as the thiostrepton inducible *rbpA* condition system in which the RbpA BL mutants were integrated (Results Chapter I).

6.3.3 Investigating a possible relationship between the DNA binding regions in RbpA and CarD

The CarD W85 is a well conserved residue across the Actinobacteria and has been predicted to interact with the open/closed promoter junction at dsDNA (-12) and ssDNA (-11). RNAP holoenzyme crystal structures have also eluded to RbpA BL interaction with the extended -10 promoter element (Hubin et al., 2017a). Both RbpA and CarD contact DNA from opposing directions and this suggested a possible overlapping role. This section focused on implementing the *sspB*-DAS dependant system to investigate a possible relationship between the two DNA binding determinants in RbpA and CarD. There is a well-defined phenotype for RbpA^{3KRA}/CarD^{WT} however the phenotype for RbpA^{3KRA}/CarD^{W85A} a lot more striking. Whilst establishing a relationship between RbpA and CarD *in vivo*, two studies were published focusing on the RbpA and CarD relationship *in vitro* (Hubin et al., 2017a; Rammohan et al., 2016). The crystal structure of RbpA/RNAP-SigA complex in *M. smegmatis* was superimposed with the CarD/RPo structure in *T. aquaticus*, which revealed that both transcription factors interact with the -10 promoter element from opposite sides of the DNA (Hubin et al., 2017a). Abortive initiation assays investigating the transcription activation at the VapB and AP3 promoters showed that the presence of RbpA and CarD increased transcription greater than RbpA and CarD alone (Hubin et al., 2017a). In addition to this, the strong -35 element of AP3 was substituted with the weak -35 element of *vapB* to create AP3^{anti-35}, transcription activation was increased with both RbpA and CarD present and appears to compensate for a weak -35 element (Hubin et al., 2017a). The binding kinetics of RbpA and CarD appear to increase the rate and amount of RPo formation and stabilise RPo (Hubin et al., 2017a; Rammohan et al., 2016).

General discussion

7.0 Overview

RbpA was first discovered when it consistently co-eluted with *Streptomyces* RNAP during affinity chromatography and gel filtration (Newell et al., 2006; Paget et al., 2001). RbpA was shown to confer basal levels of resistance to rifampicin in the first instance, and able to activate transcription *in vitro* (Tabib-Salazar et al., 2013). The $\Delta rbpA$ mutants in *S. coelicolor* produced a small colony and slow growing phenotype but were still viable, however *rbpA* and *rbpB* knockouts were not viable (Aline Tabib-Salazar and Mark Paget, personal communication) and *rbpB* was found to be non-essential. This demonstrated that RbpA together with RbpB conferred an essential function in *S. coelicolor* (Newell et al., 2006), whilst it was shown that RbpA is essential in *M. tuberculosis* with no known equivalent of RbpB (Forti et al., 2011). Additionally, RbpA was shown to activate transcription from σ^{HrdB} -dependent promoters in *S. coelicolor* (Tabib-Salazar et al., 2013) and σ^A -dependant promoters in *M. tuberculosis* (Hu et al., 2012). Later on, kinetic studies proved that RbpA was responsible for stabilisation of the RNAP holoenzyme at the transcription initiation bubble (Bortoluzzi et al., 2013; Hubin et al., 2017; Prusa et al., 2018; Rammohan et al., 2016).

During the period of this study, multiple publications provided new insights that have improved our knowledge of the structure and function of RbpA. Firstly, the RbpA BL R79 residue and SID in *M. tuberculosis* were shown to be sufficient for transcription activation and interactions with promoter DNA (Hubin et al., 2015). Secondly, several solved crystal structures provided evidence of the close interaction of the RbpA BL to the promoter DNA upstream of the extended -10 promoter element (Boyaci et al., 2018; Hubin et al., 2017; Hubin et al., 2015). Thirdly, cryo-EM techniques revealed that the RbpA NTT was positioned to thread through the active cleft of RNAP, raising important questions of the interaction between rifampicin and RbpA (Boyaci et al., 2018). Additionally, knock-outs of RbpA NTT have demonstrated that it is not required for viability in *M. smegmatis*, but does cause transcriptional changes identified using RNA-seq (Hubin et al., 2017; Prusa et al., 2018).

7.1 Highlights of this study

The first aim of this study was to investigate the importance of the RbpA BL in *S. coelicolor*. This study has shown that the RbpA BL in *S. coelicolor* is important for growth although not essential

as it grows more than the empty vector control in each case. The RbpA BL is also responsible for producing significantly more ACT than RbpA^{WT}. It appears that the lysine residues are not important alone, however may carry some overlapping function with R80. Transcriptional analysis of the RbpA BL mutant has demonstrated a global transcriptional change, this could be a direct effect or the lack of a functional RbpA could inhibit growth thereby causing an indirect global transcriptional down-regulation of most genes.

Secondly, this study aimed at investigating the importance of the RbpA BL in *M. tuberculosis*. The RbpA BL mutants did not exhibit any growth-related phenotype in liquid culture, however a growth phenotype was clear on solid media. We did see a difference in culturability for all of the RbpA BL mutants, using the MPN assay. The MPN assay paired with CFUs indicates that the empty vector and RbpA^{3KRA} are able to grow in liquid to higher dilutions compared to solid media. This has suggested that the PI inducible system may be leaking expression of *rbpA* in the absence of the inducer suggestive or that the bacilli are entering a non-culturable state. The RbpA^{3KRA} transcriptome was investigated in further detail using RNA-seq technology, this showed down-regulation of the devR mediated stress response.

Thirdly, this study investigated the mechanism and biological importance of RbpA in *S. coelicolor*. The well conserved M85 residue appears to fulfil an important role in sigma binding, amino acid substitutions demonstrate a distinct small colony phenotype with ACT production. Multiple crystals structures have found that the RbpA-SID interacts with the σ_2 domain of the group I sigma factor (Hubin et al., 2017; Hubin et al., 2015; Tabib-Salazar et al., 2013). This study has shown that substituting well conserved residues in σ^{HrdB} , important for RbpA binding to the corresponding residues of a group II does not prevent binding to σ^{HrdB} -dependent genes. This was the first study to show that HrdB^{4xR} can still associate with RNAP albeit at a lower level than HrdB^{WT}. This suggests that RbpA is not required for RNAP binding and promoter interactions *in vivo*. The HrdB^{4xR} was mapped to most of the HrdB^{WT} binding sites across the *S. coelicolor* genome but showed a reduced level of binding. Further to this, transcriptional analysis of a $\Delta rbpA$ null mutant strain demonstrated global transcriptional changes, a large proportion of this is down-regulation of growth-related pathways and induction of stress responses.

Like RbpA, CarD is an essential transcription factor in *M. tuberculosis*, $\Delta carD$ mutants exhibited a slow growth phenotype (Stallings et al., 2009b). CarD is up-regulated in response to DSBs and

nutrient limitation (Stallings et al., 2009b). ChIP-seq experiments showed that HrdB co-localises with RbpA and CarD at HrdB-dependant promoters (Laurence Humphrey, personal communication). This was supported by structural models at the time which superimposed RbpA and CarD contacting the upstream of the extended -10 promoter DNA from opposing directions. This preliminary evidence strongly suggested a potential relationship between the two essential transcription factors. To test a possible relationship between RbpA and CarD, we engineered a DAS-tagged system to exploit the cells protein degradation pathway for controlled co-depletion of *rbpA* and *carD*. The *carD* depletion did not work as clearly as the *rbpA*-DAS depletion possibly because both transcription factors are involved in different steps in transcription.

7.2 Future directions

7.2.1 Understanding promoter structure which governs transcription initiation by RbpA and CarD

The promoter sequence determines the level of expression of a gene and is recognised as a sequence 10-35 bases upstream of the transcriptional start point whereby RNAP docks and enables transcription of a gene (Newton-Foot & Pittius, 2013). Generally *E. coli* promoters start with a purine base at position +1 whilst *M. tuberculosis* TSS do not favour any particular base at +1, possibly because of the difference in GC content (Newton-Foot & Pittius, 2013). The -35 sequence is found to be more elusive for most promoters possibly because of the variety of sigma factors that can recognise a single promoter (Newton-Foot & Pittius, 2013). Additionally, the extended -10 promoter element (TGn) plays an important role in determining promoter strength in mycobacteria. Understanding the differences in promoter structures for both *M. tuberculosis* and *S. coelicolor* compared to *E. coli* could provide a better insight into the dependency of certain promoters on transcriptional regulators.

Genes belonging to a ribosomal RNA operon have been previously tested by Laurence Humphrey (unpublished data), *rrnp3* which consists of a 'poor' -35 and 'good' -10 and was shown to be dependent on RbpA and/or CarD, whilst *rrnp2* was constitutively expressed in the absence of RbpA/CarD and consisted of a 'good' -10 and 'good' -35 promoter element. The p2 'good' -35 was swapped with the corresponding 'poor' -35 from p3, this resulted in a poor level

of transcription activation at the p2 promoter. Additionally, a FLAG-tagged HrdB^{4xR} mutant that cannot bind to RbpA has shown less binding at the promoters of genes on the periphery of the *S. coelicolor* genome compared to HrdB^{WT}. The sequence alignments of the promoters for these genes, has shown conservation of the consensus -10 whilst no real consensus -35 was recognised. This further indicates that genes regulated by RbpA are composed of a poor -35 consensus sequence in *S. coelicolor*. To follow up these experiments and to establish a link between the presence of a poor -35 promoter element and dependency on RbpA, in vitro transcription assays and S1 nuclease mapping techniques are required using the RbpA^{DAS+4}/CarD^{DAS+4} system.

Additionally, a recent publication investigated the importance of the recognition of the extended -10 DNA element for transcription initiation. RbpA was shown to be dispensable for promoters which consist of an extended -10 promoter element (TG motif), whilst necessary for RPo stabilisation (Perumal et al., 2018)

7.2.2 Exploiting cryptic gene cluster for antibiotic production

Thus far, this project has shown that manipulation of the RbpA basic linker results in production of ACT. This could be a result of the down-regulation of growth related genes and hence RNAP is free to bind to other factors involved in secondary metabolite production. Several studies have shown RbpA closely interacts with the upstream extended -10 promoter region, -13/-14, (Boyaci et al., 2018; Hubin et al., 2015, 2017) so there is a potential that RbpA prevents antibiotic regulatory proteins binding to promoters involved in antibiotic production. The RbpA BL mutants could be exploited for production of antibiotic gene clusters, achieved by over-expression of the RbpA^{3KRA} allele which consists of a strong promoter and ribosome binding site. This RbpA^{3KRA} mutant allele expression should outcompete native RbpA for σ^{HrdB} and thus would mimic the BL mutant phenotype. Alternatively, random mutagenesis of RbpA could help isolate mutants that are responsible for production of ACT.

We have identified that RbpA and CarD co-localise at the same HrdB-dependant promoters (Aline Tabib-Salazar, Personal communication) during log-phase growth. To analyse these transcription factors during log-phase and late-stationary phase growth using ChIP-seq may provide an insight into binding level changes of RbpA and CarD during the switch to antibiotic

production in late-stationary phase. This PhD project has attempted to substitute the native *rbpA* promoter with the *pstS* promoter which is highly expressed under phosphate limiting conditions. This *ppstS-rbpA* mutant however did not overproduce ACT therefore data was not shown in this project (Appendix 4).

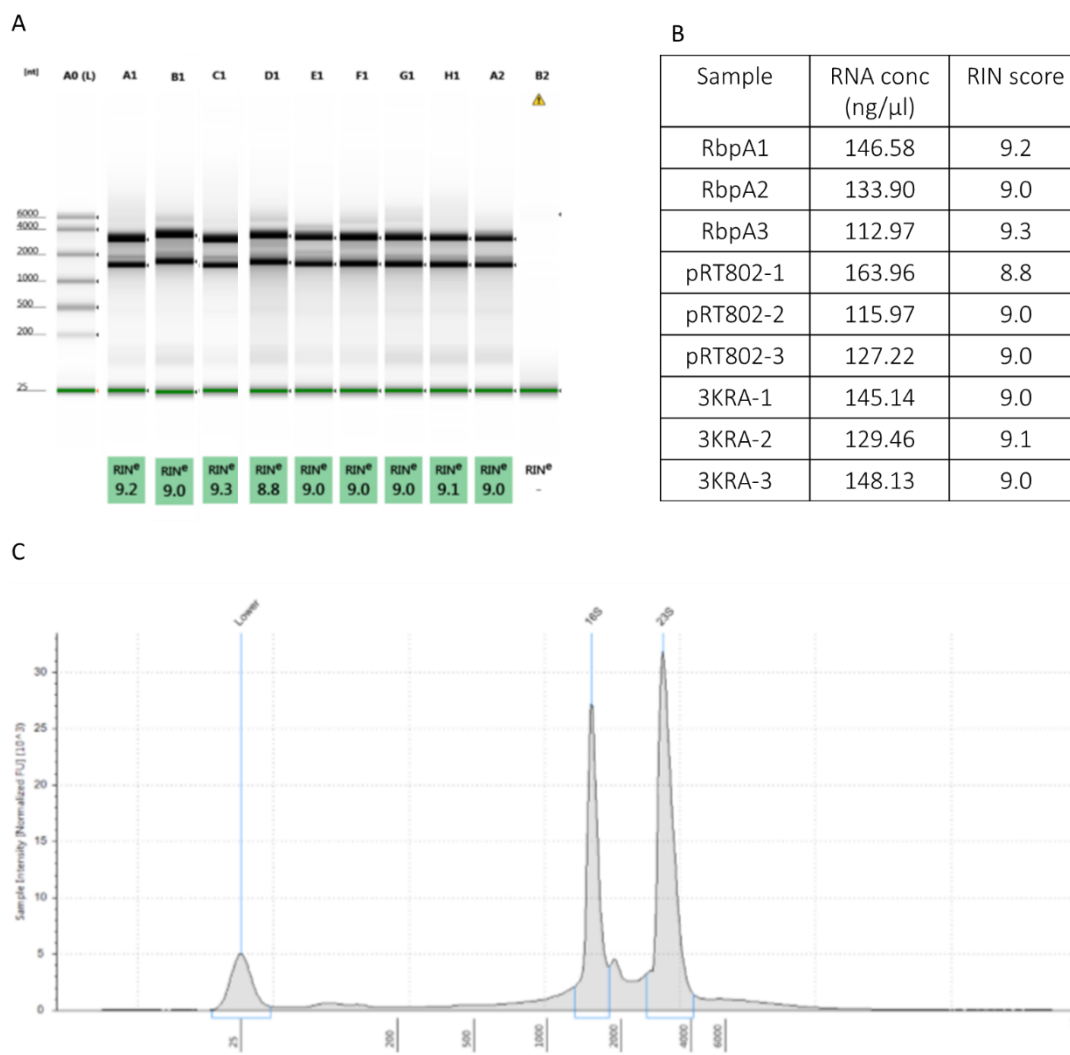
Synthetic biology for the discovery of novel antibacterial compounds is a rapidly expanding area. More specifically, the genomics era has allowed the utilisation of Actinobacteria for the activation of 'cryptic gene clusters' *in vitro* for isolation of antibacterial compounds. Most Actinobacteria only produce a few secondary metabolites under standard laboratory conditions whilst harbouring 30-40 secondary metabolite gene clusters in their genome (Ahmed et al., 2017). One such approach to activate these 'cryptic gene clusters', is the use of cell free *in vitro* transcription-translation system developed in *E. coli* and has since been applied in Actinobacteria *spp* (Moore et al., 2017). The advances of next generation sequencing and better databases for cataloguing gene clusters using for example, antiSMASH and CLUSTERtools helps drive advancement (Rutledge & Challis, 2015; Santos & Challis, 2017).

Additionally, the research focus of many Actinobacteria groups has been aimed at isolating novel bacteria that are not culturable under standard laboratory conditions. Methods to alleviate this problem have been developed, for example iChIP (isolation ChIP) which enables *in-situ* isolation of non-culturable bacteria in the native environment. This approach uses a semi-permeable membranes to isolate the organisms and the device is placed back into the soil which promotes cultivation of organisms in the native environment (Rutledge & Challis, 2015).

Appendices

Appendix 1

Quality control results from Oxford genomics centre for the RNA-sequencing samples



Agilent 4200 Tapestation quality control results for the *S. coelicolor* BL RNA samples from Oxford genomics centre. A. Gel image of the RNA profiles, lanes A1, B1 and C1 correspond to the 3 biological replicates for RbpA^{WT}, lanes D1-F1 correspond to the 3 biological replicates for pRT802, and lanes G1-A2 correspond to the 3 biological replicates of RbpA^{3KRA}. B. RNA concentrations and RNA integrity number equivalent (RIN) of the RNA samples. C. RNA profile of S401 (pRT802::*rbpA*^{WT}) (1) showing the ribosomal subunits 16S and 23S.

Appendix 2

Purification of core RNAP:

Optimisation of growth conditions for RNAP purification

Several conditions have been tested initially to maximise the production of soluble core RNA polymerase (Fig. A1). However, attempts to purify native RNAP from *S. coelicolor* was unsuccessful, therefore recombinant RNAP in *E. coli* was used optimised and utilised instead. Firstly, Rosetta 2 (DE3) pLysS cells were transformed with pSU904E (created by Simon Hodges), single colonies were used to inoculate cultures overnight. The culture was diluted to 1:100 and grown to OD₆₀₀ 0.8, followed by testing multiple variables; 1 mM or 250 µM IPTG, with or without cold shock, 16°C or 20°C and 4 h or 24 h post IPTG induction. 2mL samples pre/post IPTG induction were collected and cells were lysed by sonication 6 times for 10 sec with 30 sec on/off ice after each cycle at 35%. Soluble fractions were run on 10% SDS-PAGE gels.

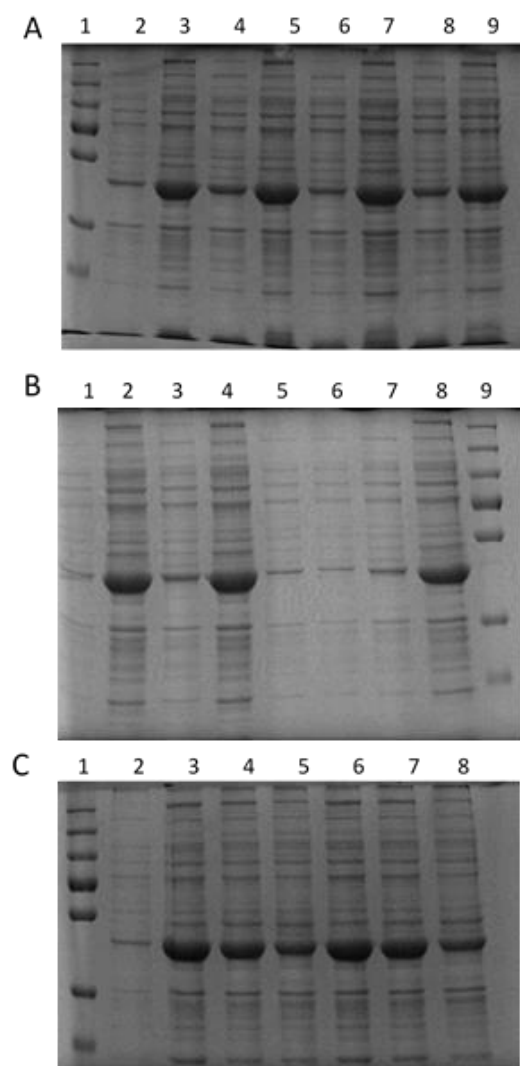


Figure A1. Optimisation of pSU904E recombinant core RNA polymerase in *E. coli*. SDS-PAGE analysis of soluble recombinant core RNA polymerase purification. (A) 1- PageRuler Prestained protein marker, 2- 4 h post 1 mM IPTG at 20°C, 3- overnight post 1 mM at 20°C, 4- 4 h post 1 mM IPTG at 20°C without cold shock, 5- overnight post 1 mM at 20°C without cold shock, 6- 4 h post 250 μ M IPTG at 20°C, 7- overnight post 250 μ M IPTG at 20°C, 8- 4 h post 250 μ M IPTG at 20°C without cold shock, 9- overnight post 250 μ M at 20°C without cold shock. (B) 1- 4 h post 1 mM IPTG at 16°C, 2- overnight post 1 mM at 16°C, 3- 4 h post 1 mM IPTG at 16°C without cold shock, 4- overnight post 1 mM at 16°C without cold shock, 5- 4 h post 250 μ M IPTG at 16°C, 6- overnight post 250 μ M at 16°C, 7- 4 h post 250 μ M IPTG at 16°C without cold shock, 8- overnight post 250 μ M at 16°C without cold shock, 9- PageRuler Prestained protein marker. (C) 1- PageRuler Prestained protein marker, 2- Pre-IPTG addition soluble fraction, 3- 24 h 1 mM IPTG, 4- 48 h 1 mM IPTG, 5- 72 h 1 mM IPTG, 6- 24 h 250 μ M IPTG, 7- 48 h 250 μ M IPTG, 8- 72 h 250 μ M IPTG.

The recombinant *S. coelicolor* RNA polymerase was overexpressed from plasmid pSU904E in *E. coli* BL21 λ DE3 pRARE2. RNAP overexpression was induced by the addition of 500 μ M IPTG to cultures at OD₆₀₀ 0.8 and the incubation was continued for 24 h at 20°C without any cold shock prior to adding the IPTG. The cells were recovered by centrifugation and stored at -80°C. The cell pellet from 500 mL cultures was re-suspended in 30 mL 20 mM Tris-HCl pH 7.9, 5 mM imidazole, 500 mM NaCl, 5mM β - mercaptoethanol, 1 mM fresh PMSF. The cell suspension was lysed by sonication 10 times with 30 sec ON and OFF at 35% and centrifuged. The cleared cell lysate was bound to a 5mL Nickel Sepharose HP HisTrap column (GE healthcare). The column was washed with 20 mM Tris-HCl pH 7.9, 5 mM imidazole, 500 mM NaCl, 5 mM β -mercaptoethanol until no A280 was detectable, and eluted with a 40 mL gradient of 5 to 2000 mM imidazole in 20 mM Tris-HCl pH 7.9, 500 mM NaCl, 5mM β -mercaptoethanol. The pooled fractions containing RNAP were dialyzed into 10 mM Tris-HCl pH 7.9; 200 mM NaCl, 0.1 mM EDTA, 1 mM MgCl₂, 20 μ M ZnCl₂, 10 mM DTT however RNAP had crashed out of solution and a precipitation was visible. The presence of EDTA and MgCl₂ in the same buffer was responsible for the precipitation. Therefore, the purification was repeated by using a combination of different gel filtration buffers:

Buffer A: 10 mM Tris-HCl pH 7.9, 500 mM NaCl, 5% glycerol, 10 mM DTT

Buffer B: 10 mM Tris-HCl pH 7.9, 500 mM NaCl, 0.1mM EDTA, 10 mM DTT

Buffer C: 10 mM Tris-HCl pH 7.9, 500 mM NaCl, 5% glycerol, 1mM MgCl₂, 20 μ M ZnCl₂, 10 mM DTT

Fractions were run on a 10% SDS-PAGE gel and silver stained to observe the best purification after size exclusion and dialysis. Buffer C provided the best purification and therefore a large scale purification was undertaken to purify core RNA polymerase.

3L core Rosetta II BL21 (pSU904E) chemically competent cells were set up at 37 °C to OD₆₀₀ 0.8 and induced with 500 μ M IPTG and further cultured for 24 h at 20 °C. The 3L pellet was re-suspended and sonicated to produce a cleared cell lysate which was loaded onto the 5 mL HisTrap Nickel column using the Akta Prime. A 2 M imidazole elution gradient was used to produce fractions (Fig. A2A). Fractions around 600 mM imidazole were pooled together and dialysed overnight using buffer C (Fig. A2B). The pooled sample was concentrated to 2 mL using a 15 mL Vivaspin device (10,000 MW cut off; GE Healthcare) following the manufacturer's

instructions. The sample was loaded onto a pre-calibrated (Fig. A2C) the size exclusion Superdex 200 column (GE healthcare). Fractions containing RNA polymerase were pooled (Fig A2D), concentrated to ≥ 1 mg RNAP/mL, dialyzed overnight at 4 °C with storage buffer, and stored in aliquots at -80 °C (Fig A2E).

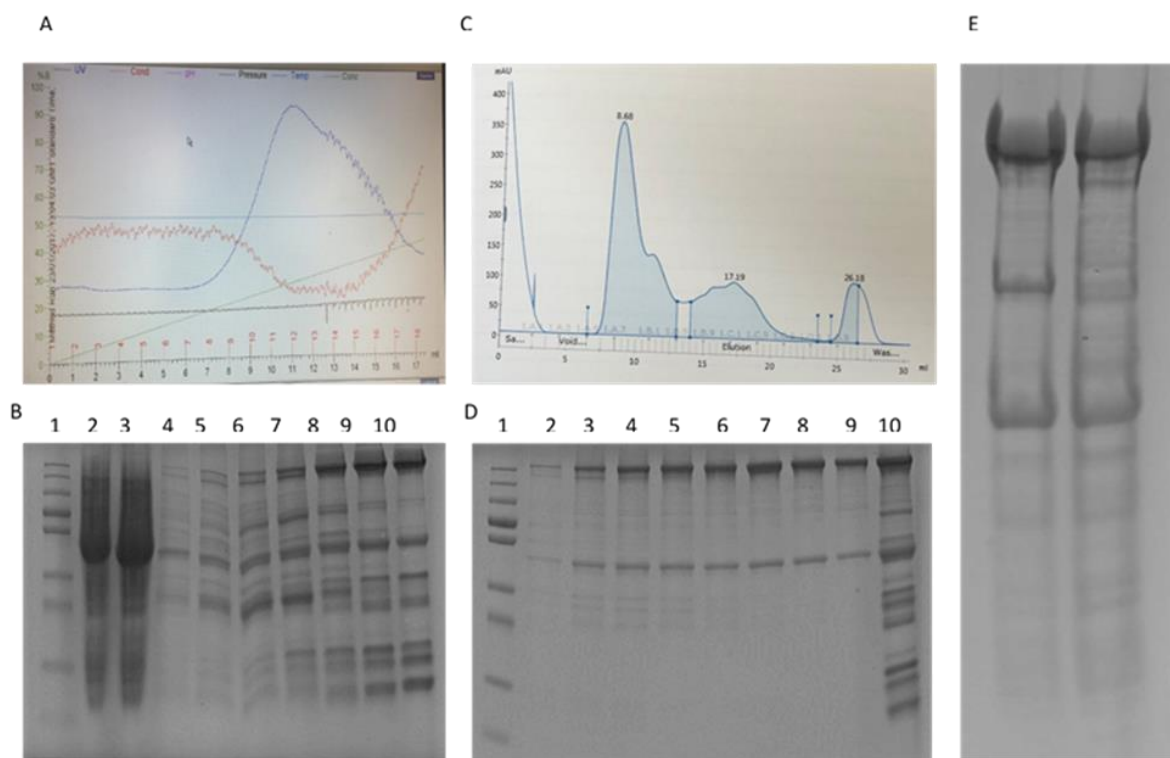


Figure A2. Purification of recombinant *S. coelicolor* RNA polymerase. Rosetta II BL21 chemically competent cells transformed with pSU904E containing the *S. coelicolor* recombinant core RNA polymerase subunits. 3L LB cultures for Rosetta II BL21 with pSU904E cultured at 37 °C to OD₆₀₀ 0.8 and induced with 500 µM IPTG and continued culturing at 20 °C for 24 h. The cultures were harvested by centrifugation 4 °C, the cell pellet was re-suspended in 60 mL 20 mM Tris-HCl pH 7.9, 5 mM imidazole, 500 mM NaCl, 5mM β-mercaptoethanol, 1 mM PMSF. The resuspension was sonicated for 10 cycles of 30 sec ON/OFF at 35% frequency. The lysate was cleared by centrifugation and passed through a 0.2 µm filter. The cleared lysate was loaded onto a 5 mL nickel HisTrap column following the manufacturer's instructions using the Akta Prime. (A) UV reading from the Akta prime for the fractionated elutions using a 2 M imidazole gradient concentration. (B) 1- cell lysate, 2-flow through, 3-9 fractions from the HisTrap using 2 M imidazole gradient elution with the Akta prime. The protein fractions 3-9 were pooled together and dialysed with buffer C overnight. The dialysed sample was centrifuged at full speed for 10 min at 4 °C before loading onto the Size exclusion Superdex 200 column. (C) UV read out of the size exclusion column showing three peaks. The (D) fractions corresponding to the peaks were run on a 10% SDS-PAGE and stained with Coomassie brilliant blue stain, 1- PageRuler Pre-stained marker, 2-9 fractions, 10- post dialysis overnight sample before size exclusion column. The fractions from the size exclusion column were pooled together and concentrated using a

Vivaspin column to achieve 500 µL sample. The sample was dialysed overnight using the storage buffer, 10 mM Tris-HCl pH7.9, 500 mM NaCl, 1 mM MgCl₂, 20 µM ZnCl₂, 50% glycerol and 2 mM DTT at 4 °C. (E) Purified core RNA polymerase samples post dialysis and stored in storage buffer at -80 °C.

Appendix 3

Investigating the dependency of RbpA and CarD on promoters

Previous studies, indicate that RbpA and CarD may compensate for the loss of a -35 consensus sequence at certain promoters (Hubin et al., 2017). Therefore, this section focused on investigating the *in vitro* transcription of RbpA and CarD at the *S. coelicolor* ribosomal *rrnD* operon, *rrnDp2* and *rrnDp3*, promoters. The *rrnDp2* promoter consists of a strong -10 and -35 consensus sequences whilst the *rrnDp3* consists of a strong -10 and weak -35 consensus sequences. The pSRE3 plasmid (created by Sven Reisloehner) is a derivative of pRT802 and consists of *ypet* (yellow protein) and kanamycin resistance marker. The pSRE3 plasmid was modified by Sven Reisloehner so that the *ypet* protein does not have a start codon or ribosome binding site, resulting in the inhibition of translation and production of unstable mRNA which was ideal for this experiment as the level of expression at the promoters was investigated using S1 nuclease mapping. To investigate the dependency of RbpA and CarD on *rrnDp2* and *rrnDp3* and on the -35 element, promoters were designed by swapping the consensus -35 sequence from *rrnDp2* with the poor consensus sequence of *rrnDp3*, this has been referred to as *rrnDp2-3* (Fig. A3). The poor -35 element from *rrnDp3* was also swapped with the consensus -35 sequence from *rrnDp2*, this is referred to as *rrnDp3-2* (Fig. A3). The oligos were annealed together and inserted into pSRE3 using the BamHI and HindIII sites (Table 2.8).

rrnDp2 *GATCCTTCCCGCAAGAGCCGTTGACACGGAGCGAGCGGGGAGGTAGATT**CGAACAGTA*
 GAAGGGCGTTCTCGGCAACTGTGCCTCGCTCGCCCCCTCCATCTAAGCTTGTCA*TTCGA*

rrnDp3 *GATCCCGGAAACGAAGGCCGGTAAGACCGGCTCGAAAGTTCTGATAAAAGTCGGAGCCGCA*
 GGCCTTTGCTTCCGGCCATTCTGGCCGAGCTTTCAAGACTATTTTCAGCCTCGGCG*TTCGA*

rrnDp2-3 *GATCCTTCCCGCAAGAGCCGTAAGACCGGAGCGAGCGGGGAGGTAGATT**CGAACAGTA*
 GAAGGGCGTTCTCGGCATTCTGGCCTCGCTCGCCCCCTCCATCTAAGCTTGTCA*TTCGA*

rrnDp3-2 *GATCCCGGAAACGAAGGCCGGTTGACACGGCTCGAAAGTTCTGATAAAAGTCGGAGCCGCA*
 GGCCTTTGCTTCCGGCCAACTGTGCCGAGCTTTCAAGACTATTTTCAGCCTCGGCG*TTCGA*

Figure. A3 Promoters tested for dependency on RbpA and CarD. The *rrnDp2* good -35 consensus sequence was swapped with the *rrnDp3* poor -35 consensus sequence to create *rrnDp2-3*. The *rrnDp3* poor -35 consensus sequence was swapped with the *rrnDp2* good -35 consensus sequence which created *rrnDp3-2*. The restriction sites are shown, BamHI (in italics) and HindIII (underlined and italics) and the TSS is shown in red. The TSS for the *rrnD* operon were identified by Baylis and Bibb (1988).

All of the promoters were conjugated into the S416 (pIJ6902::*sspB*) strain. To optimise the S1 nuclease mapping experiment, only the S416 (pIJ6902::*sspB*; pSRE3::*p2-3*) plasmid was cultured in liquid media. The cells from this strain were subjected to thiostrepton treatment for RbpA and CarD co-depletion, followed by RNA extractions at 0, 2, 4 and 6 h. The S1 oligo was designed to overlap the *ypet* gene with an 8 bp random sequence at the 3' (Fig. A4B) and radiolabelled with ³²P on the 5' end. To perform the S1 nuclease mapping, 30 µg of RNA for each time point was mixed with 5-10 ng of probe, samples were run on an 8% denaturing urea polyacrylamide gel and visualised by phosphorimaging. Unfortunately, none of the samples showed a signal for any transcript in any of the wells. The whole S1 nuclease mapping experiment was repeated for the S416 (pIJ6902::*sspB*; pSRE3::*p3*) and the same results were obtained (Fig. A5). This suggests that there was a problem with the construction of the plasmid.

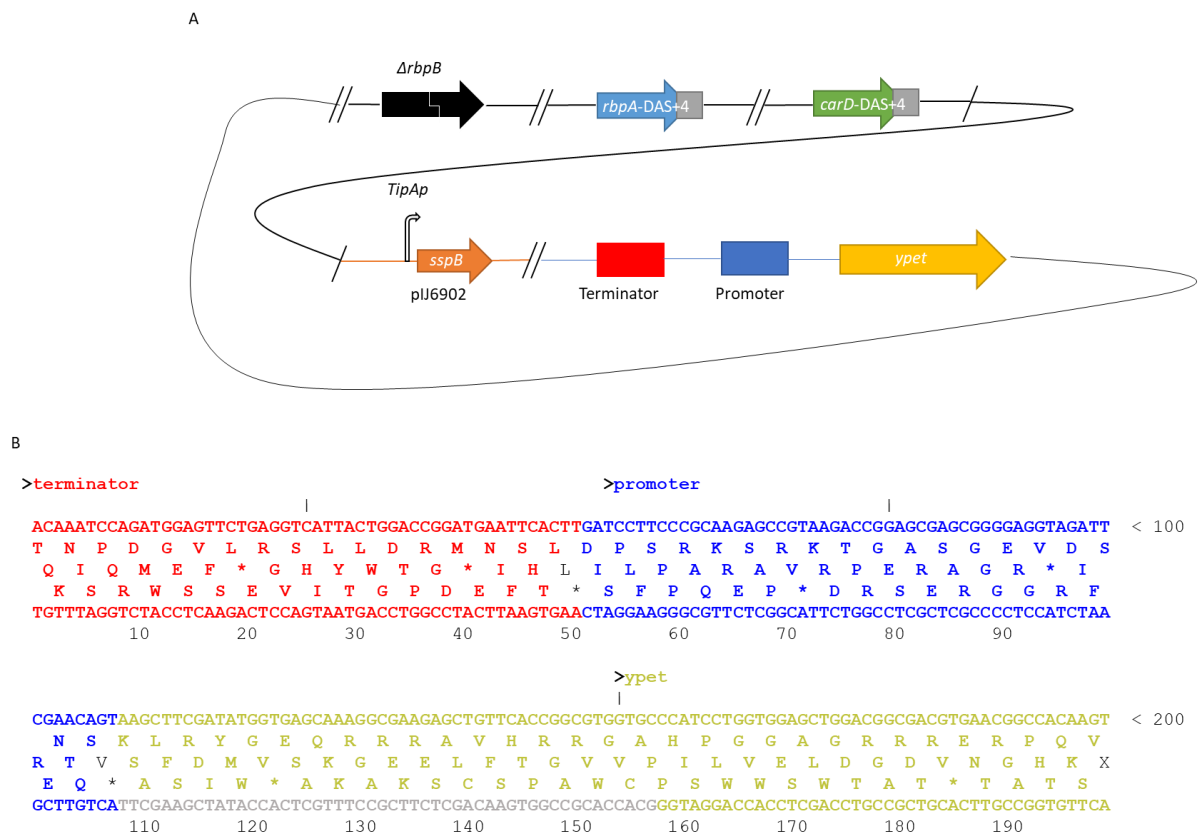


Figure. A4 Organisation of the S416 RbpA^{DAS+4}/CarD^{DAS+4} pSRE3::p2-3 (pIJ6902::sspB) strain. (A) The strain consists of $\Delta rbpB$, *rbpA*-DAS+4/*carD*-DAS+4 with pIJ6902::sspB integrated at the phiC31 site, *sspB* is under the control of a thiostrepton *tipA* promoter. pSRE3 has integrated at the phiBT1 *attP* site and the p2-3 promoter has been inserted as a BamHI-HindIII fragment into pSRE3, upstream of the *ypet* gene and downstream of the terminator. (B) The terminator (red), rrnDp2-3 promoter (blue) and *ypet* gene (yellow). The S1 nuclease mapping oligo (grey) aligns from a 3'- 5' direction with RNA in a 5'-3' direction. The oligo consists of 8 bp random nucleotides 3' GTGGCTGT.

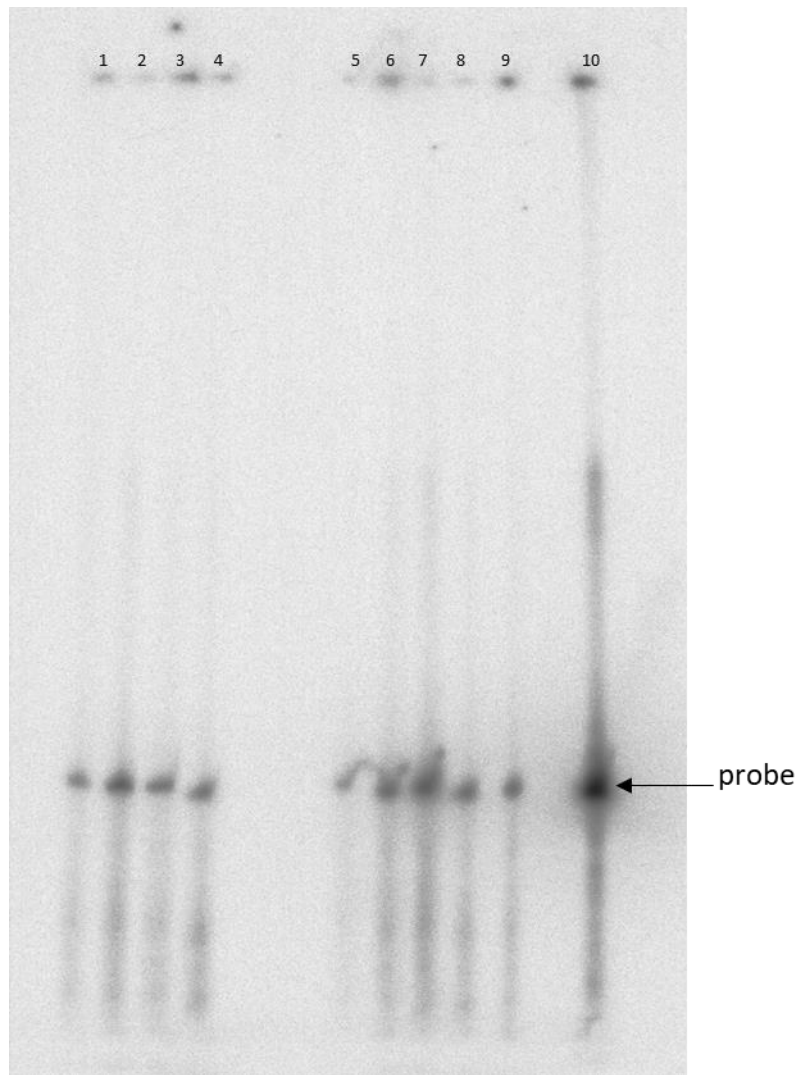


Figure A5. S1 nuclease mapping of the S416 (pIJ6902::*sspB*; pSRE3::*p2-3*) \pm thiostrepton. No thiostrepton treatment was used as a control in lane 1-5. Lane 1- 0 h, Lane 2- 2 h, Lane 3- 4 h, Lane 4- 6 h. Samples in lanes 6-8 contain 15 µg/mL thiostrepton, Lane 5- 0 h, Lane 6- 2 h, Lane 7- 4 h, Lane 8- 6 h. Lanes 9 and 10 contain the probe.

Appendix 4

Manipulation of RbpA to over-produce actinorhodin

Streptomyces has long been used to find novel compounds suitable for therapeutic purposes. *S. coelicolor* produces up to 12 useful compounds, of these is actinorhodin (ACT). ACT is overproduced in the RbpA^{3KRA} mutants and therefore this section aims to identify promoters that are strongly expressed which could be used for the overexpression of RbpA^{WT} and RbpA^{3KRA} via a fusion of the strong promoter and translational signals to RbpA. Transcriptional start site data showed that *psts* (SCO4142) was highly expressed during transition, late exponential and stationary phase (Jeong et al, 2016). TSS RNA-seq data from *Streptomyces venezuelae* (SVEN15_3808) showed that *psts* is switched on at transition phase when the cells are depleted of phosphate, and highly expressed throughout transition and late log phase (Matt Hutchings, personal communication). The sequence alignment of SVEN *psts* against SCO *psts* showed that the transcriptional and translational signals are highly conserved. *pstS* is switched on upon low phosphate conditions which is detected by the two component regulatory system *phoR-phoP*. PhoP recognises PHO boxes in the regulon, one of which is *psts*. The *psts* promoter was translationally fused to the ORF RbpA with the 3XFLAG tag using a two-stage PCR method. Figure A6 shows the experimental set up of the two stage PCR for the fusion of the *psts* promoter to *rbpA* 3XFLAG tag (Table 10).

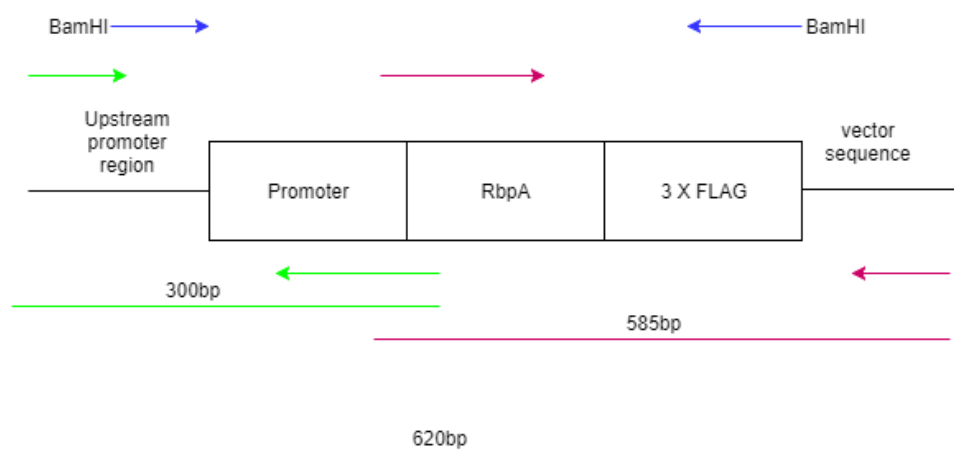


Figure A6. Two stage PCR for the over-expression of RbpA. Oligos in green are for the upstream PCR product, oligos in magenta are for downstream PCR product and the oligos in blue amplify the promoter, RbpA and 3 X FLAG in a single PCR product using the upstream and downstream PCR products from stage 1 as the template for the second stage.

The first stage of the two-stage PCR involves using the oligos Psts UF and Psts UR to amplify the upstream Psts fragment using *S. coelicolor* genome DNA whilst the downstream (Psts DF and Psts DR) fragment was amplified using pRT802::*rbpA*^{WT}3XFLAG or pRT802::*rbpA*^{3KRA}3XFLAG as the template. The second stage of the PCR was set up using the *psts* upstream PCR product and *psts* downstream PCR product. The second stage of the PCR uses the oligos Psts UBamHI and Psts DBamHI (Table 2.14). This produced one single fragment ~620 bp which was gel purified. PCR products from the second stage PCR were ligated to pBluescript using the EcoRV restriction site, blue white screening was used to screen for recombinants and subsequent recombinants were confirmed with sequencing. The confirmed insert was digested with BamHI and ligated into the integrative vectors, pRT802 and pSET152. Recombinants were screened and the subsequent recombinants for pRT802 were transformed into an ETR *E. coli* conjugative strain, whilst the pSET152 were transformed into ETZ *E. coli*. These were finally conjugated into the M145 (wild-type *S. coelicolor*), S101 (Δ *rbpA*) and S401 (J1915 Δ *rbpA* Δ *rbpB*) backgrounds. The samples were plated onto 0-0.5 mM phosphate concentrations which demonstrated no increased production of ACT in the absence of *rbpA* (Fig A7) nor was there any increased production of ACT in the S401 background between *RbpA*^{WT} and *RbpA*^{3KRA}.



Figure. A7 Actinorhodin over-production in an M145 and S101 *Streptomyces* background. The M145 pRT802::*pstS-rbpA*^{WT}, M145 pRT802::*pstS-rbpA*^{3KRA}, M145 pRT802, S101 pRT802::*pstS-rbpA*^{WT}, S101 pRT802::*pstS-rbpA*^{3KRA}, S101 pRT802 were plated onto SMMS agar at 0-0.5 mM phosphate concentrations and cultured for 3-4 days at 30°C.

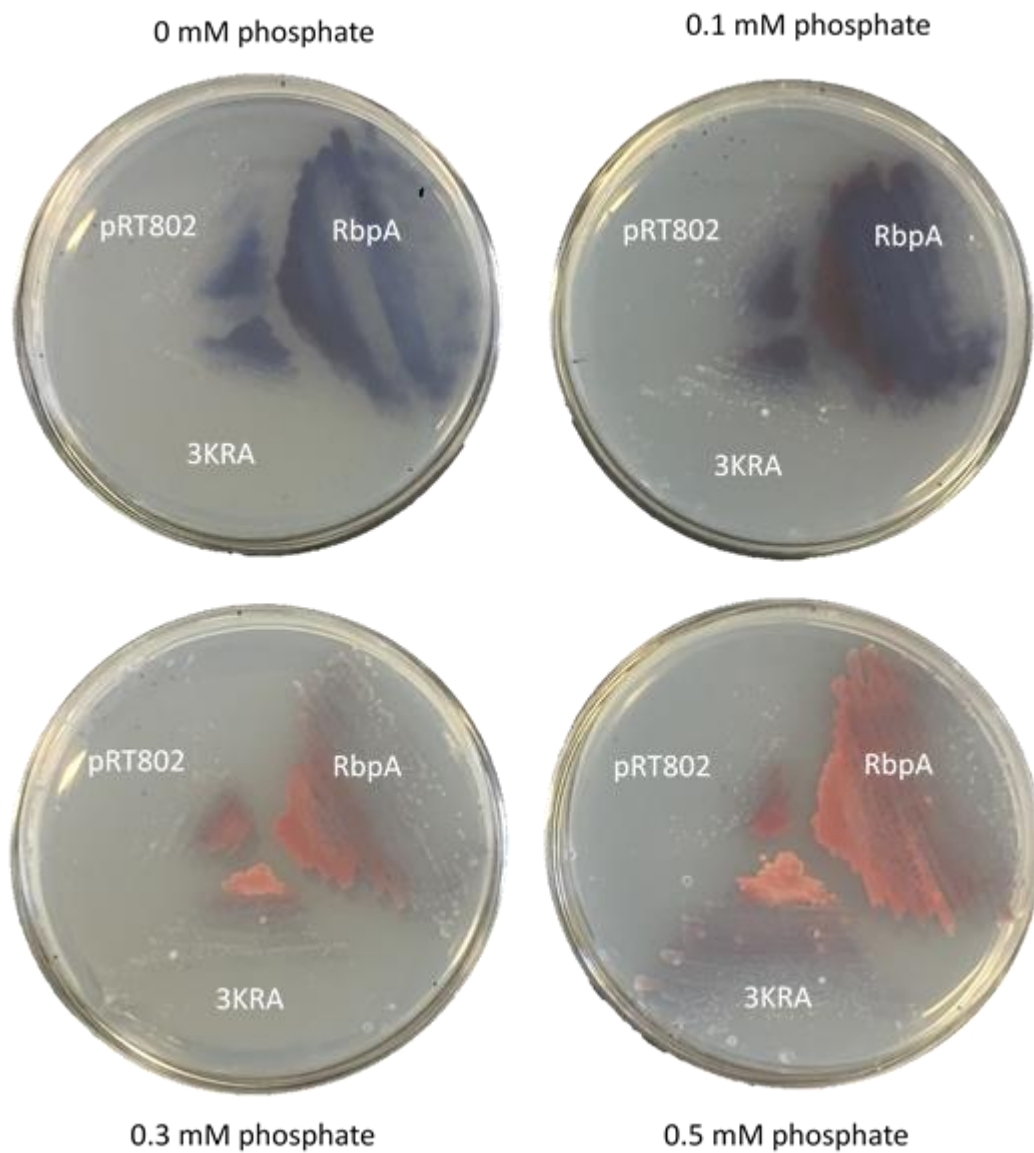


Figure. A8 Actinorhodin over-production in the S401 genetic background. S401 (pRT802), S401 (pRT802::*pstS-rbpA^{WT}*) and S401 (pRT802::*pstS-rbpA^{3KRA}*) were plated onto SMMS in the absence of thiostrepton and phosphate concentrations 0-0.5 mM. The plates were cultured at 30°C for 3-4 days.

References

- Abreu, C., Rocha-Pereira, N., Sarmiento, A., & Magro, F. (2015). Nocardia infections among immunomodulated inflammatory bowel disease patients: A review. *World Journal of Gastroenterology : WJG*, 21(21), 6491–6498. <https://doi.org/10.3748/wjg.v21.i21.6491>
- Agarwal, N., & Tyagi, A. K. (2006). Mycobacterial transcriptional signals: Requirements for recognition by RNA polymerase and optimal transcriptional activity. *Nucleic Acids Research*, 34(15), 4245–4257. <https://doi.org/10.1093/nar/gkl521>
- Ahmed, Y., Rebets, Y., Tokovenko, B., Brötz, E., & Luzhetskyy, A. (2017). Identification of butenolide regulatory system controlling secondary metabolism in *Streptomyces albus* J1074. *Scientific Reports*, 7(1). <https://doi.org/10.1038/s41598-017-10316-y>
- Archambault, J., & Friesen, J. D. (1993). Genetics of eukaryotic RNA polymerases I, II, and III. *Microbiological Reviews*, 57, 703–724.
- Arnqvist, A., Olsén, A., Pfeifer, J., Russell, D. G., & Normark, S. (1992). The Crl protein activates cryptic genes for curli formation and fibronectin binding in *Escherichia coli* HB101. *Molecular Microbiology*, 6(17), 2443–2452. <https://doi.org/10.1111/j.1365-2958.1992.tb01420.x>
- Babcock, M. J., Buttner, M. J., Keler, C. H., Clarke, B. R., Morris, R. A., Lewis, C. G., & Brawner, M. E. (1997). Characterization of the *rpoC* gene of *Streptomyces coelicolor* A3(2) and its use to develop a simple and rapid method for the purification of RNA polymerase. *Gene*, 196(1–2), 31–42. [https://doi.org/10.1016/S0378-1119\(97\)00179-0](https://doi.org/10.1016/S0378-1119(97)00179-0)
- Bae, B., Chen, J., Davis, E., Leon, K., Darst, S. A., & Campbell, E. A. (2015). CarD uses a minor groove wedge mechanism to stabilize the RNA polymerase open promoter complex. *ELife*, 4(September 2015). <https://doi.org/10.7554/eLife.08505>
- Bardou, F., Raynaud, C., Ramos, C., Lanéelle, M. A., & Lanéelle, G. (1998). Mechanism of isoniazid uptake in *Mycobacterium tuberculosis*. *Microbiology*, 144(9), 2539–2544. <https://doi.org/10.1099/00221287-144-9-2539>
- Barka, E. A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Klenk, H.-P., ... van Wezel, G. P. (2016). Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiology and Molecular Biology Reviews*, 80(1), 1–43. <https://doi.org/10.1128/MMBR.00019-15>
- Baron, S. (1996). *Medical Microbiology. 4th edition. University of Texas Medical Branch at Galveston*. <https://doi.org/NBK8035> [bookaccession]
- Bashyam, M. D., Kaushal, D., Dasgupta, S. K., & Tyagi, A. K. (1996). A study of the mycobacterial transcriptional apparatus: Identification of novel features in promoter elements. *Journal of Bacteriology*, 178(16), 4847–4853. <https://doi.org/10.1128/jb.178.16.4847-4853.1996>

- Baylis, H. A., & Bibb, M. J. (1988). Transcriptional analysis of the 16S rRNA gene of the *rrnD* gene set of *Streptomyces coelicolor* A3(2). *Mol Microbiol*, 2(5), 569–579. <https://doi.org/10.1111/j.1365-2958.1988.tb00065.x>
- Béahdy, J. (1974). Recent Developments of Antibiotic Research and Classification of Antibiotics According to Chemical Structure. *Advances in Applied Microbiology*, 18(C), 309–406. [https://doi.org/10.1016/S0065-2164\(08\)70573-2](https://doi.org/10.1016/S0065-2164(08)70573-2)
- Bear, D. G., Hicks, P. S., Escudero, K. W., Andrews, C. L., McSwiggen, J. A., & von Hippel, P. H. (1988). Interactions of *Escherichia coli* transcription termination factor rho with RNA. II. Electron microscopy and nuclease protection experiments. *Journal of Molecular Biology*, 199(4), 623–635. [https://doi.org/10.1016/0022-2836\(88\)90306-3](https://doi.org/10.1016/0022-2836(88)90306-3)
- Becker, N. A., Peters, J. P., & Maher, L. J. (2013). Mechanism of promoter repression by Lac repressor-DNA loops. *Nucleic Acids Research*, 41(1), 156–166. <https://doi.org/10.1093/nar/gks1011>
- Bentley, S. D., Chater, K. F., Cerdeño-Tárraga, A.-M., Challis, G. L., Thomson, N. R., James, K. D., ... Hopwood, D. A. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature*, 417, 141–147. <https://doi.org/10.1038/417141a>
- Bermudez, L. E., & Goodman, J. (1996). *Mycobacterium tuberculosis* invades and replicates within type II alveolar cells. *Infection and Immunity*, 64(4), 1400–1406.
- Betts, J. C., Lukey, P. T., Robb, L. C., McAdam, R. A., & Duncan, K. (2002). Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. *Molecular Microbiology*, 43(3), 717–731. <https://doi.org/10.1046/j.1365-2958.2002.02779.x>
- Bierman, M., Logan, R., O'Brien, K., Seno, E. T., Nagaraja Rao, R., & Schoner, B. E. (1992). Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene*, 116(1), 43–49. [https://doi.org/10.1016/0378-1119\(92\)90627-2](https://doi.org/10.1016/0378-1119(92)90627-2)
- Borovok, I., Gorovitz, B., Schreiber, R., Aharonowitz, Y., & Cohen, G. (2006). Coenzyme B12 controls transcription of the *Streptomyces* class Ia ribonucleotide reductase *nrdABS* operon via a riboswitch mechanism. *Journal of Bacteriology*, 188(7), 2512–2520. <https://doi.org/10.1128/JB.188.7.2512-2520.2006>
- Borovok, I., Gorovitz, B., Yanku, M., Schreiber, R., Gust, B., Chater, K., ... Cohen, G. (2004). Alternative oxygen-dependent and oxygen-independent ribonucleotide reductases in *Streptomyces*: Cross-regulation and physiological role in response to oxygen limitation. *Molecular Microbiology*, 54(4), 1022–1035. <https://doi.org/10.1111/j.1365-2958.2004.04325.x>
- Borovok, I., Kreisberg-Zakarin, R., Yanko, M., Schreiber, R., Myslovati, M., Åslund, F., ... Aharonowitz, Y. (2002). *Streptomyces* spp. contain class Ia and class II ribonucleotide reductases: Expression analysis of the genes in vegetative growth. *Microbiology*, 148(2), 391–404. <https://doi.org/10.1099/00221287-148-2-391>

- Bortoluzzi, A., Muskett, F. W., Waters, L. C., Addis, P. W., Rieck, B., Munder, T., ... O'Hare, H. M. (2013). Mycobacterium tuberculosis RNA polymerase-binding protein A (RbpA) and its interactions with sigma factors. *Journal of Biological Chemistry*, 288, 14438–14450. <https://doi.org/10.1074/jbc.M113.459883>
- Borukhov, S., Lee, J., & Laptenko, O. (2005). Bacterial transcription elongation factors: New insights into molecular mechanism of action. *Molecular Microbiology*. <https://doi.org/10.1111/j.1365-2958.2004.04481.x>
- Botella, L., Vaubourgeix, J., Livny, J., & Schnappinger, D. (2017). Depleting Mycobacterium tuberculosis of the transcription termination factor Rho causes pervasive transcription and rapid death. *Nature Communications*, 8. <https://doi.org/10.1038/ncomms14731>
- Boyaci, H., Chen, J., Lilic, M., Palka, M., Mooney, R. A., Landick, R., ... Campbell, E. A. (2018). Fidaxomicin jams mycobacterium tuberculosis RNA polymerase motions needed for initiation via RBPA contacts. *ELife*, 7. <https://doi.org/10.7554/eLife.34823>
- Brennan, M. J. (2017). The enigmatic PE/PPE multigene family of mycobacteria and tuberculosis vaccination. *Infection and Immunity*. <https://doi.org/10.1128/IAI.00969-16>
- Brosch, R., Gordon, S. V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., ... Cole, S. T. (2002). A new evolutionary scenario for the Mycobacterium tuberculosis complex. *Proceedings of the National Academy of Sciences*, 99(6), 3684–3689. <https://doi.org/10.1073/pnas.052548299>
- Burgess, R. R., Travers, a a, Dunn, J. J., & Bautz, E. K. (1969). Factor stimulating transcription by RNA polymerase. *Nature*, 221, 43–46. <https://doi.org/10.1038/221043a0>
- Bush, M. J., Bibb, M. J., Chandra, G., Findlay, K. C., & Buttner, M. J. (2013). Genes required for aerial growth, cell division, and chromosome segregation are targets of whia before sporulation in Streptomyces venezuelae. *MBio*, 4(5). <https://doi.org/10.1128/mBio.00684-13>
- Buttner, M. J., Chater, K. F., & Bibb, M. J. (1990). Cloning, disruption, and transcriptional analysis of three RNA polymerase sigma factor genes of Streptomyces coelicolor A3(2). *Journal of Bacteriology*, 172(6), 3367–3378.
- Buttner, M. J., Smith, A. M., & Bibb, M. J. (1988). At least three different RNA polymerase holoenzymes direct transcription of the agarase gene (dagA) of streptomyces coelicolor A3(2). *Cell*, 52(4), 599–607. [https://doi.org/10.1016/0092-8674\(88\)90472-2](https://doi.org/10.1016/0092-8674(88)90472-2)
- Bystrykh, L. V., Fernández-Moreno, M. A., Herrema, J. K., Malpartida, F., Hopwood, D. A., & Dijkhuizen, L. (1996). Production of actinorhodin-related “blue pigments” by Streptomyces coelicolor A3(2). *Journal of Bacteriology*, 178(8), 2238–2244. <https://doi.org/10.1128/JB.178.8.2238-2244.1996>

- Campagne, S., Marsh, M. E., Capitani, G., Vorholt, J. A., & Allain, F. H. T. (2014). Structural basis for -10 promoter element melting by environmentally induced sigma factors. *Nature Structural and Molecular Biology*, 21(3), 269–276. <https://doi.org/10.1038/nsmb.2777>
- Campbell, E. A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., & Darst, S. A. (2001). Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell*, 104, 901–912. [https://doi.org/10.1016/S0092-8674\(01\)00286-0](https://doi.org/10.1016/S0092-8674(01)00286-0)
- Campbell, E. a, Pavlova, O., Zenkin, N., Leon, F., Irschik, H., Jansen, R., ... Darst, S. a. (2005). Structural, functional, and genetic analysis of sorangicin inhibition of bacterial RNA polymerase. *The EMBO Journal*, 24(4), 674–682. <https://doi.org/10.1038/sj.emboj.7600499>
- Capstick, D. S., Willey, J. M., Buttner, M. J., & Elliot, M. A. (2007). SapB and the chaplins: Connections between morphogenetic proteins in *Streptomyces coelicolor*. *Molecular Microbiology*, 64(3), 602–613. <https://doi.org/10.1111/j.1365-2958.2007.05674.x>
- Chakraborty, R., & Bibb, M. (1997). The ppGpp synthetase gene (*relA*) of *Streptomyces coelicolor* A3(2) plays a conditional role in antibiotic production and morphological differentiation. *Journal of Bacteriology*, 179(18), 5854–5861. <https://doi.org/10.1128/jb.179.18.5854-5861.1997>
- CHATER, K. F. (1972). A Morphological and Genetic Mapping Study of White Colony Mutants of *Streptomyces coelicolor*. *Journal of General Microbiology*, 72(1), 9–28. <https://doi.org/10.1099/00221287-72-1-9>
- Chater, K. F. (1975). Construction and Phenotypes of Double Sporulation Deficient Mutants in *Streptomyces coelicolor* A3(2). *Journal of General Microbiology*, 87(2), 312–325. <https://doi.org/10.1099/00221287-87-2-312>
- Chater, K. F. (2016). Recent advances in understanding *Streptomyces*. *F1000Research*, 5(0), 2795. <https://doi.org/10.12688/f1000research.9534.1>
- Chater, K. F., & Chandra, G. (2008). The use of the rare UUA codon to define “expression space” for genes involved in secondary metabolism, development and environmental adaptation in *Streptomyces*. *Journal of Microbiology*. <https://doi.org/10.1007/s12275-007-0233-1>
- Chater, K. F., & Chandra, G. (2006). The evolution of development in *Streptomyces* analysed by genome comparisons. *FEMS Microbiology Reviews*. <https://doi.org/10.1111/j.1574-6976.2006.00033.x>
- Chauhan, R., Ravi, J., Datta, P., Chen, T., Schnappinger, D., Bassler, K. E., ... Gennaro, M. L. (2016). Reconstruction and topological characterization of the sigma factor regulatory network of *Mycobacterium tuberculosis*. *Nature Communications*, 7. <https://doi.org/10.1038/ncomms11062>

- Chen, J. M., Ren, H., Shaw, J. E., Wang, Y. J., Li, M., Leung, A. S., ... Liu, J. (2008). Lsr2 of *Mycobacterium tuberculosis* is a DNA-bridging protein. *Nucleic Acids Research*, 36(7), 2123–2135. <https://doi.org/10.1093/nar/gkm1162>
- Čihák, M., Kameník, Z., Šmídová, K., Bergman, N., Benada, O., Kofroňová, O., ... Bobek, J. (2017). Secondary Metabolites Produced during the Germination of *Streptomyces coelicolor*. *Frontiers in Microbiology*, 8, 2495. <https://doi.org/10.3389/fmicb.2017.02495>
- Claessen, D., Rink, R., De Jong, W., Siebring, J., De Vreugd, P., Boersma, F. G. H., ... Wösten, H. A. B. (2003). A novel class of secreted hydrophobic proteins is involved in aerial hyphae formation in *Streptomyces coelicolor* by forming amyloid-like fibrils. *Genes and Development*, 17(14), 1714–1726. <https://doi.org/10.1101/gad.264303>
- Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., ... Barrell, B. G. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*. <https://doi.org/10.1038/31159>
- Cortes, T., & Cox, R. A. (2015). Transcription and translation of the *rpsJ*, *rplN* and *rRNA* operons of the tubercle bacillus. *Microbiology (United Kingdom)*, 161(4), 719–728. <https://doi.org/10.1099/mic.0.000037>
- Cortes, T., Schubert, O. T., Rose, G., Arnvig, K. B., Comas, I., Aebersold, R., & Young, D. B. (2013). Genome-wide Mapping of Transcriptional Start Sites Defines an Extensive Leaderless Transcriptome in *Mycobacterium tuberculosis*. *Cell Reports*, 5(4), 1121–1131. <https://doi.org/10.1016/j.celrep.2013.10.031>
- Croucher, N. J., & Thomson, N. R. (2010). Studying bacterial transcriptomes using RNA-seq. *Current Opinion in Microbiology*. <https://doi.org/10.1016/j.mib.2010.09.009>
- Czyz, A., Mooney, R. A., Iaconi, A., & Landick, R. (2014). Mycobacterial RNA polymerase requires a U-tract at intrinsic terminators and is aided by NusG at suboptimal terminators. *MBio*, 5(2). <https://doi.org/10.1128/mBio.00931-14>
- Darst, S. A., Kubalek, E. W., & Kornberg, R. D. (1989). Three-dimensional structure of *Escherichia coli* RNA polymerase holoenzyme determined by electron crystallography. *Nature*, 340(6236), 730–732. <https://doi.org/10.1038/340730a0>
- Dasgupta, N., Kapur, V., Singh, K. K., Das, T. K., Sachdeva, S., Jyothisri, K., & Tyagi, J. S. (2000). Characterization of a two-component system, *devR-devS*, of *Mycobacterium tuberculosis*. *Tubercle and Lung Disease*, 80(3), 141–159. <https://doi.org/10.1054/tuld.2000.0240>
- Davis, E., Chen, J., Leon, K., Darst, S. A., & Campbell, E. A. (2015). Mycobacterial RNA polymerase forms unstable open promoter complexes that are stabilized by CarD. *Nucleic Acids Research*, 43(1), 433–445. <https://doi.org/10.1093/nar/gku1231>
- Dawes, S. S., Warner, D. F., Tsenova, L., Timm, J., McKinney, J. D., Kaplan, G., ... Mizrahi, V. (2003). Ribonucleotide Reduction in *Mycobacterium tuberculosis*: Function and

- Expression of Genes Encoding Class Ib and Class II Ribonucleotide Reductases. *Infection and Immunity*, 71(11), 6124–6131. <https://doi.org/10.1128/IAI.71.11.6124-6131.2003>
- Deng, L., Mikusova, K., Robuck, K. G., Scherman, M., Brennan, P. J., & McNeil, M. R. (1995). Recognition of multiple effects of ethambutol on metabolism of mycobacterial cell envelope. *Antimicrobial Agents and Chemotherapy*, 39(3), 694–701. <https://doi.org/10.1128/AAC.39.3.694>
- Dey, A., Verma, A. K., & Chatterji, D. (2010). Role of an RNA polymerase interacting protein, MsRbpA, from *Mycobacterium smegmatis* in phenotypic tolerance to rifampicin. *Microbiology*, 156(3), 873–883. <https://doi.org/10.1099/mic.0.033670-0>
- Dombroski, A. J., Walter, W. A., & Gross, C. A. (1993). Amino-terminal amino acids modulate sigma-factor DNA-binding activity. *Genes & Development*, 7(12A), 2446–2455. <https://doi.org/10.1101/gad.7.12a.2446>
- Doroghazi, J. R., & Metcalf, W. W. (2013). Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes. *BMC Genomics*, 14, 611. <https://doi.org/10.1186/1471-2164-14-611>
- Doukhan, L., Predich, M., Nair, G., Dussurget, O., Mandic-Mulec, I., Cole, S. T., ... Smith, I. (1995). Genomic organization of the mycobacterial sigma gene cluster. *Gene*, 165(1), 67–70. [https://doi.org/10.1016/0378-1119\(95\)00427-8](https://doi.org/10.1016/0378-1119(95)00427-8)
- Dye, C. (2013). Making wider use of the world's most widely used vaccine: Bacille Calmette-Guérin revaccination reconsidered. *Journal of the Royal Society Interface*. <https://doi.org/10.1098/rsif.2013.0365>
- Ebright, R. (1995). The *Escherichia coli* RNA polymerase alpha subunit: structure and function. *Current Opinion in Genetics & Development*, 5(2), 197–203. [https://doi.org/10.1016/0959-437X\(95\)80008-5](https://doi.org/10.1016/0959-437X(95)80008-5)
- Ehrt, S., & Schnappinger, D. (2014). Regulated Expression Systems for *Mycobacteria* and Their Applications. *Microbiology Spectrum*, 2(1). <https://doi.org/10.1128/microbiolspec.MGM2-0018-2013>
- Estrem, S. T., Ross, W., Gaal, T., Chen, Z. W. S., Niu, W., Ebright, R. H., & Gourse, R. L. (1999). Bacterial promoter architecture: Subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase α subunit. *Genes and Development*, 13(16), 2134–2147. <https://doi.org/10.1101/gad.13.16.2134>
- Falkinham, J. O. (2009). Surrounded by mycobacteria: Nontuberculous mycobacteria in the human environment. *Journal of Applied Microbiology*. <https://doi.org/10.1111/j.1365-2672.2009.04161.x>
- Feklistov, A. (2013). RNA polymerase: In search of promoters. *Annals of the New York Academy of Sciences*, 1293(1), 25–32. <https://doi.org/10.1111/nyas.12197>

- Feklistov, A., & Darst, S. A. (2011). Structural basis for promoter -10 element recognition by the bacterial RNA polymerase σ subunit. *Cell*, 147(6), 1257–1269. <https://doi.org/10.1016/j.cell.2011.10.041>
- Fernández-Martínez, L. T., & Bibb, M. J. (2014). Use of the Meganuclease I-SceI of *Saccharomyces cerevisiae* to select for gene deletions in actinomycetes. *Scientific Reports*, 4. <https://doi.org/10.1038/srep07100>
- Fink, D., Weissschuh, N., Reuther, J., Wohlleben, W., & Engels, a. (2002). Two transcriptional regulators GlnR and GlnRII are involved in regulation of nitrogen metabolism in *Streptomyces coelicolor* A3(2). *Molecular Microbiology*, 46(2), 331–47. <https://doi.org/3150> [pii]
- Fisher, S. H., & Wray, L. V. (1989). Regulation of glutamine synthetase in *Streptomyces coelicolor*. *Journal of Bacteriology*, 171(5), 2378–2383. <https://doi.org/10.1128/jb.171.5.2378-2383.1989>
- Flärdh, K. (2003). Growth polarity and cell division in *Streptomyces*. *Current Opinion in Microbiology*. <https://doi.org/10.1016/j.mib.2003.10.011>
- Flärdh, K. (2003). Essential role of DivIVA in polar growth and morphogenesis in *Streptomyces coelicolor* A3(2). *Molecular Microbiology*, 49(6), 1523–1536. <https://doi.org/10.1046/j.1365-2958.2003.03660.x>
- Flärdh, K., & Buttner, M. J. (2009). *Streptomyces* morphogenetics: Dissecting differentiation in a filamentous bacterium. *Nature Reviews Microbiology*. <https://doi.org/10.1038/nrmicro1968>
- Flentje, K., Garner, A. L., & Stallings, C. L. (2016). Mycobacterium tuberculosis transcription machinery: Ready to respond to host attacks. *Journal of Bacteriology*. <https://doi.org/10.1128/JB.00935-15>
- Flynn, J. M., Levchenko, I., Seidel, M., Wickner, S. H., Sauer, R. T., & Baker, T. a. (2001). Overlapping recognition determinants within the *ssrA* degradation tag allow modulation of proteolysis. *Proceedings of the National Academy of Sciences of the United States of America*, 98(19), 10584–10589. <https://doi.org/10.1073/pnas.191375298>
- Flynn, J. M., Levchenko, I., Sauer, R. T., & Baker, T. A. (2004). Modulating substrate choice: The SspB adaptor delivers a regulator of the extracytoplasmic-stress response to the AAA+ protease ClpXP for degradation. *Genes and Development*, 18(18), 2292–2301. <https://doi.org/10.1101/gad.1240104>
- Flynn, J. M., Neher, S. B., Kim, Y. I., Sauer, R. T., & Baker, T. A. (2003). Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Molecular Cell*, 11(3), 671–683. [https://doi.org/10.1016/S1097-2765\(03\)00060-1](https://doi.org/10.1016/S1097-2765(03)00060-1)
- Forbes, B. A. (2017). Mycobacterial taxonomy. *Journal of Clinical Microbiology*, 55(2), 380–383. <https://doi.org/10.1128/JCM.01287-16>

- Forrellad, M. A., Klepp, L. I., Gioffré, A., Sabio y García, J., Morbidoni, H. R., de la Paz Santangelo, M., ... Bigi, F. (2013). Virulence factors of the *Mycobacterium tuberculosis* complex. *Virulence*, 4(1), 3–66. <https://doi.org/10.4161/viru.22329>
- Forti, F., Mauri, V., Deh, G., & Ghisotti, D. (2011). Isolation of conditional expression mutants in *Mycobacterium tuberculosis* by transposon mutagenesis. *Tuberculosis*, 91(6), 569–578. <https://doi.org/10.1016/j.tube.2011.07.004>
- Gaal, T., Mandel, M. J., Silhavy, T. J., & Gourse, R. L. (2006). Crl facilitates RNA polymerase holoenzyme formation. *Journal of Bacteriology*, 188(22), 7966–7970. <https://doi.org/10.1128/JB.01266-06>
- Gao, B., & Gupta, R. S. (2012). Phylogenetic Framework and Molecular Signatures for the Main Clades of the Phylum Actinobacteria. *Microbiology and Molecular Biology Reviews*, 76(1), 66–112. <https://doi.org/10.1128/MMBR.05011-11>
- Garner, A. L., Rammohan, J., Huynh, J. P., Onder, L. M., Chen, J., Bae, B., ... Stallings, C. L. (2017). Effects of increasing the affinity of CarD for RNA polymerase on *Mycobacterium tuberculosis* growth, rRNA transcription, and virulence. *Journal of Bacteriology*, 199(4). <https://doi.org/10.1128/JB.00698-16>
- Garner, A. L., Weiss, L. a, Manzano, A. R., Galburt, E. a, & Stallings, C. L. (2014). CarD integrates three functional modules to promote efficient transcription, antibiotic tolerance, and pathogenesis in mycobacteria. *Molecular Microbiology*, 93(4), 682–97. <https://doi.org/10.1111/mmi.12681>
- Gatfield, J., & Pieters, J. (2000). Essential role for cholesterol in entry of mycobacteria into macrophages. *Science*, 288(5471), 1647–1650. <https://doi.org/10.1126/science.288.5471.1647>
- Gomez, M., Doukhan, L., Nair, G., & Smith, I. (1998). SigA is an essential gene in *Mycobacterium smegmatis*. *Molecular Microbiology*, 29(2), 617–628. <https://doi.org/10.1046/j.1365-2958.1998.00960.x>
- Gomez, M., Doukhan, L., Nair, G., & Smith, I. (1998). SigA is an essential gene in *Mycobacterium smegmatis*. *Molecular Microbiology*, 29(2), 617–628. <https://doi.org/10.1046/j.1365-2958.1998.00960.x>
- Gomez-Escribano, J. P., Song, L., Fox, D. J., Yeo, V., Bibb, M. J., & Challis, G. L. (2012). Structure and biosynthesis of the unusual polyketide alkaloid coelimycin P1, a metabolic product of the cpk gene cluster of *Streptomyces coelicolor* M145. *Chemical Science*, 3(9), 2716. <https://doi.org/10.1039/c2sc20410j>
- Gordon, B. R. G., Li, Y., Wang, L., Sintsova, A., van Bakel, H., Tian, S., ... Liu, J. (2010). Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences*, 107(11), 5154–5159. <https://doi.org/10.1073/pnas.0913551107>

- Gordon, B. R. G., Imperial, R., Wang, L., Navarre, W. W., & Liu, J. (2008). Lsr2 of *Mycobacterium* represents a novel class of H-NS-like proteins. *Journal of Bacteriology*, 190(21), 7052–7059. <https://doi.org/10.1128/JB.00733-08>
- Gottesman, S., Roche, E., Zhou, Y., & Sauer, R. T. (1998). The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes and Development*, 12(9), 1338–1347. <https://doi.org/10.1101/gad.12.9.1338>
- Gregory, M. A., Till, R., & Smith, M. C. M. (2003). Integration site for *Streptomyces* phage ϕ BT1 and development of site-specific integrating vectors. *Journal of Bacteriology*, 185(17), 5320–5323. <https://doi.org/10.1128/JB.185.17.5320-5323.2003>
- Gries, T. J., Kontur, W. S., Capp, M. W., Saecker, R. M., & Record, M. T. (2010). One-step DNA melting in the RNA polymerase cleft opens the initiation bubble to form an unstable open complex. *Proceedings of the National Academy of Sciences of the United States of America*, 107(23), 10418–23. <https://doi.org/10.1073/pnas.1000967107>
- Griffith, K. L., & Grossman, A. D. (2008). Inducible protein degradation in *Bacillus subtilis* using heterologous peptide tags and adaptor proteins to target substrates to the protease ClpXP. *Molecular Microbiology*, 70(4), 1012–1025. <https://doi.org/10.1111/j.1365-2958.2008.06467.x>
- Guirado, E., & Schlesinger, L. S. (2013). Modeling the *Mycobacterium tuberculosis* granuloma - the critical battlefield in host immunity and disease. *Frontiers in Immunology*, 4(APR). <https://doi.org/10.3389/fimmu.2013.00098>
- Haneishi, T., Terahara, A., Arai, M., Hata, T., & Tamura, C. (1974). New antibiotics, Methylenomycins A AND B. <https://doi.org/http://doi.org/10.7164/antibiotics.27.393>
- Hartmann, G., Honikel, K. O., Knüsel, F., & Nüesch, J. (1967). The specific inhibition of the DNA-directed RNA synthesis by rifamycin. *BBA Section Nucleic Acids And Protein Synthesis*, 145(3), 843–844. [https://doi.org/10.1016/0005-2787\(67\)90147-5](https://doi.org/10.1016/0005-2787(67)90147-5)
- Heep, M., Rieger, U., Beck, D., & Lehn, N. (2000). Mutations in the beginning of the *rpoB* gene can induce resistance to rifamycins in both *Helicobacter pylori* and *Mycobacterium tuberculosis*. *Antimicrobial Agents and Chemotherapy*, 44(4), 1075–7. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=89817&tool=pmcentrez&rendertype=abstract>
- Hesketh, A., Chen, W. J., Ryding, J., Chang, S., & Bibb, M. (2007). The global role of ppGpp synthesis in morphological differentiation and antibiotic production in *Streptomyces coelicolor* A3(2). *Genome Biology*, 8(8). <https://doi.org/10.1186/gb-2007-8-8-r161>
- Hett, E. C., & Rubin, E. J. (2008). Bacterial growth and cell division: a mycobacterial perspective. *Microbiology and Molecular Biology Reviews : MMBR*, 72(1), 126–56, table of contents. <https://doi.org/10.1128/MMBR.00028-07>

- Homerova, D., Halgasova, L., & Kormanec, J. (2008). Cascade of extracytoplasmic function sigma factors in *Mycobacterium tuberculosis*: Identification of a σ J-dependent promoter upstream of sigI. *FEMS Microbiology Letters*, 280(1), 120–126. <https://doi.org/10.1111/j.1574-6968.2007.01054.x>
- Hong, H. J., Paget, M. S. B., & Buttner, M. J. (2002). A signal transduction system in *Streptomyces coelicolor* that activates the expression of a putative cell wall glycan operon in response to vancomycin and other cell wall-specific antibiotics. *Molecular Microbiology*, 44(5), 1199–1211. <https://doi.org/10.1046/j.1365-2958.2002.02960.x>
- Hopwood, D. A., & Wright, H. M. (1983). CDA is a new chromosomally-determined antibiotic from *Streptomyces coelicolor* A3(2). *Journal of General Microbiology*. <https://doi.org/10.1099/00221287-129-12-3575>
- Hopwood, D. A., Bibb, M. J., Chater, K. F., Kieser, T., Bruton, C. J., Kieser, H. M., ... Schrempf, H. (1986). *Genetic manipulation of Streptomyces: A laboratory manual. Biochemical Education* (Vol. 14). [https://doi.org/10.1016/0307-4412\(86\)90228-1](https://doi.org/10.1016/0307-4412(86)90228-1)
- Hopwood, D., Wildermuth, H., & Palmer, H. (1970). Mutants of *Streptomyces coelicolor* Defective in Sporulation. *Journal of General Microbiology*, 61, 397–408. <https://doi.org/10.1099/00221287-61-3-397>
- Hopwood, D. A. (1999). Forty years of genetics with *Streptomyces*: From in vivo through in vitro to in silico. *Microbiology*. <https://doi.org/10.1099/00221287-145-9-2183>
- Hu, H., Zhang, Q., & Ochi, K. (2002). Activation of antibiotic biosynthesis by specified mutations in the rpoB gene (encoding the RNA polymerase β' subunit) of *Streptomyces lividans*. *Journal of Bacteriology*, 184, 3984–3991. <https://doi.org/10.1128/JB.184.14.3984-3991.2002>
- Hu, Y., Morichaud, Z., Chen, S., Leonetti, J. P., & Brodolin, K. (2012). *Mycobacterium tuberculosis* RbpA protein is a new type of transcriptional activator that stabilizes the σ A-containing RNA polymerase holoenzyme. *Nucleic Acids Research*, 40(14), 6547–6557. <https://doi.org/10.1093/nar/gks346>
- Hu, Y., Morichaud, Z., Sudalaiyadum Perumal, A., Roquet-Baneres, F., & Brodolin, K. (2014). *Mycobacterium* RbpA cooperates with the stress-response σ B subunit of RNA polymerase in promoter DNA unwinding. *Nucleic Acids Research*, 1–10. <https://doi.org/10.1093/nar/gku742>
- Hu, Y., Kendall, S., Stoker, N. G., & Coates, A. R. M. (2004). The *Mycobacterium tuberculosis* sigJ gene controls sensitivity of the bacterium to hydrogen peroxide. *FEMS Microbiology Letters*, 237(2), 415–423. <https://doi.org/10.1016/j.femsle.2004.07.005>
- Huang, J., Shi, J., Molle, V., Sohlberg, B., Weaver, D., Bibb, M. J., ... Cohen, S. N. (2005). Cross-regulation among disparate antibiotic biosynthetic pathways of *Streptomyces coelicolor*. *Molecular Microbiology*, 58(5), 1276–1287. <https://doi.org/10.1111/j.1365-2958.2005.04879.x>

- Hubin, E. A., Darst, S. A., & Campbell, E. A. (2017). Crystal structure of a *Mycobacterium smegmatis* transcription initiation complex with RbpA. *Doi.Org*. <https://doi.org/10.2210/pdb5tw1/pdb>
- Hubin, E. A., Fay, A., Xu, C., Bean, J. M., Saecker, R. M., Glickman, M. S., ... Campbell, E. A. (2017). Structure and function of the mycobacterial transcription initiation complex with the essential regulator RbpA. *ELife*, 6. <https://doi.org/10.7554/eLife.22520>
- Hubin, E. A., Lilic, M., Darst, S. A., & Campbell, E. A. (2017). Structural insights into the mycobacteria transcription initiation complex from analysis of X-ray crystal structures. *Nature Communications*, 8. <https://doi.org/10.1038/ncomms16072>
- Hubin, E. A., Tabib-Salazar, A., Humphrey, L. J., Flack, J. E., Olinares, P. D. B., Darst, S. A., ... Paget, M. S. (2015). Structural, functional, and genetic analyses of the actinobacterial transcription factor RbpA. *Proceedings of the National Academy of Sciences of the United States of America*, 112(23), 7171–7176. <https://doi.org/10.1073/pnas.1504942112>
- Hutchings, M. I., Hong, H. J., Leibovitz, E., Sutcliffe, I. C., & Buttner, M. J. (2006). The σ^E cell envelope stress response of *Streptomyces coelicolor* is influenced by a novel lipoprotein, CseA. *Journal of Bacteriology*, 188(20), 7222–7229. <https://doi.org/10.1128/JB.00818-06>
- Jakimowicz, D., Mouz, S., Zakrzewska-Czerwińska, J., & Chater, K. F. (2006). Developmental control of a parAB promoter leads to formation of sporulation-associated ParB complexes in *Streptomyces coelicolor*. In *Journal of Bacteriology* (Vol. 188, pp. 1710–1720). <https://doi.org/10.1128/JB.188.5.1710-1720.2006>
- Jang, H.-J., Nde, C., Toghrol, F., & Bentley, W. E. (2009). Global transcriptome analysis of the *Mycobacterium bovis* BCG response to sodium hypochlorite. *Applied Microbiology and Biotechnology*, 85(1), 127–140. <https://doi.org/10.1007/s00253-009-2208-0>
- Jarvis, B., Wilrich, C., & Wilrich, P. T. (2010). Reconsideration of the derivation of Most Probable Numbers, their standard deviations, confidence bounds and rarity values. *Journal of Applied Microbiology*, 109(5), 1660–1667. <https://doi.org/10.1111/j.1365-2672.2010.04792.x>
- Jeong, Y., Kim, J.-N., Kim, M. W., Bucca, G., Cho, S., Yoon, Y. J., ... Cho, B.-K. (2016). The dynamic transcriptional and translational landscape of the model antibiotic producer *Streptomyces coelicolor* A3(2). *Nature Communications*, 7, 11605. <https://doi.org/10.1038/ncomms11605>
- Jones, A. C., Gust, B., Kulik, A., Heide, L., Buttner, M. J., & Bibb, M. J. (2013). Phage P1-Derived Artificial Chromosomes Facilitate Heterologous Expression of the FK506 Gene Cluster. *PLoS ONE*, 8(7). <https://doi.org/10.1371/journal.pone.0069319>
- Kallifidas, D., Thomas, D., Doughty, P., & Paget, M. S. B. (2010). The σ^R regulon of *Streptomyces coelicolor* A3(2) reveals a key role in protein quality control during

- disulphide stress. *Microbiology*, 156(6), 1661–1672.
<https://doi.org/10.1099/mic.0.037804-0>
- Kang, J. G., Hahn, M. Y., Ishihama, A., & Roe, J. H. (1997). Identification of sigma factors for growth phase-related promoter selectivity of RNA polymerases from *Streptomyces coelicolor* A3(2). *Nucleic Acids Research*, 25(13), 2566–2573.
<https://doi.org/10.1093/nar/25.13.2566>
- Karimova, G., Pidoux, J., Ullmann, A., & Ladant, D. (1998). A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proceedings of the National Academy of Sciences*, 95(10), 5752–5756. <https://doi.org/10.1073/pnas.95.10.5752>
- Keiler, K. C., Waller, P. R. H., & Sauer, R. T. (1996). Role of a Peptide Tagging System in Degradation of Proteins Synthesized from Damaged Messenger RNA. *Science*, 271(5251), 990–993. <https://doi.org/10.1126/science.271.5251.990>
- Keiler, K. C. (2008). Biology of trans-translation. *Annual Review of Microbiology*, 62, 133–151.
<https://doi.org/10.1146/annurev.micro.62.081307.162948>
- Kelemen, G. H., Plaskitt, K. A., Lewis, C. G., Findlay, K. C., & Buttner, M. J. (1995). Deletion of DNA lying close to the glkA locus induces ectopic sporulation in *Streptomyces coelicolor* A3(2). *Molecular Microbiology*, 17(2), 221–230. https://doi.org/10.1111/j.1365-2958.1995.mmi_17020221.x
- Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F., & Hopwood, D. A. (2000). Practical *Streptomyces* Genetics. *John Innes Centre Ltd*. <https://doi.org/10.4016/28481.01>
- Kim, J. H., Wei, J. R., Wallach, J. B., Robbins, R. S., Rubin, E. J., & Schnappinger, D. (2011). Protein inactivation in mycobacteria by controlled proteolysis and its application to deplete the beta subunit of RNA polymerase. *Nucleic Acids Research*, 39(6), 2210–2220.
<https://doi.org/10.1093/nar/gkq1149>
- Kinashi, H., & Shimaji-Murayama, M. (1991). Physical characterization of SCP1, a giant linear plasmid from *Streptomyces coelicolor*. *Journal of Bacteriology*, 173(4), 1523–1529.
- Kodani, S., Hudson, M. E., Durrant, M. C., Buttner, M. J., Nodwell, J. R., & Willey, J. M. (2004). From The Cover: The SapB morphogen is a lantibiotic-like peptide derived from the product of the developmental gene ramS in *Streptomyces coelicolor*. *Proceedings of the National Academy of Sciences*, 101(31), 11448–11453.
<https://doi.org/10.1073/pnas.0404220101>
- Kulbachinskiy, A., & Mustaev, A. (2006). Region 3.2 of the σ subunit contributes to the binding of the 3'-initiating nucleotide in the RNA polymerase active center and facilitates promoter clearance during initiation. *Journal of Biological Chemistry*, 281(27), 18273–18276. <https://doi.org/10.1074/jbc.C600060200>
- Kumar, A., Toledo, J. C., Patel, R. P., Lancaster, J. R., & Steyn, A. J. C. (2007). Mycobacterium tuberculosis DosS is a redox sensor and DosT is a hypoxia sensor. *Proceedings of the*

- National Academy of Sciences*, 104(28), 11568–11573.
<https://doi.org/10.1073/pnas.0705054104>
- Kusano, S., Ding, Q., Fujita, N., & Ishihama, A. (1996). Promoter selectivity of Escherichia coli RNA polymerase Es70 and Es38 holoenzymes. Effect of DNA supercoiling. *J Biol Chem*, 271(4), 1998–2004.
- Lakey, J. H., Lea, E. J., Rudd, B. a, Wright, H. M., & Hopwood, D. a. (1983). A new channel-forming antibiotic from Streptomyces coelicolor A3(2) which requires calcium for its activity. *Journal of General Microbiology*, 129(12), 3565–73.
<https://doi.org/10.1099/00221287-129-12-3565>
- Lange, R., & Hengge-Aronis, R. (1994). The cellular concentration of the σ^{70} subunit of RNA polymerase in Escherichia coli is controlled at the levels of transcription, translation, and protein stability. *Genes and Development*, 8(13), 1600–1612.
<https://doi.org/10.1101/gad.8.13.1600>
- Lee, J. H., Ammerman, N. C., Nolan, S., Geiman, D. E., Lun, S., Guo, H., & Bishai, W. R. (2012). Isoniazid resistance without a loss of fitness in Mycobacterium tuberculosis. *Nature Communications*, 3. <https://doi.org/10.1038/ncomms1724>
- Leskiw, B. K., Lawlor, E. J., Fernandez-Abalos, J. M., & Chater, K. F. (1991). TTA codons in some genes prevent their expression in a class of developmental, antibiotic-negative, Streptomyces mutants. *Proceedings of the National Academy of Sciences of the United States of America*, 88(6), 2461–5. <https://doi.org/10.1073/pnas.88.6.2461>
- Liu, B., Zuo, Y., & Steitz, T. A. (2015). Structural basis for transcription reactivation by RapA. *Proceedings of the National Academy of Sciences*, 112(7), 2006–2010.
<https://doi.org/10.1073/pnas.1417152112>
- Liu, G., Chater, K. F., Chandra, G., Niu, G., & Tan, H. (2013). Molecular regulation of antibiotic biosynthesis in streptomyces. *Microbiology and Molecular Biology Reviews : MMBR*, 77(1), 112–43. <https://doi.org/10.1128/MMBR.00054-12>
- Lloyd, G., Landini, P., & Busby, S. (2001). Activation and repression of transcription initiation in bacteria. *Essays in Biochemistry*, 37, 17–31. <https://doi.org/10.1042/bse0370017>
- Lonetto, M., Gribskov, M., & Gross, C. A. (1992). The sigma 70 family: sequence conservation and evolutionary relationships. *Journal of Bacteriology*, 174(12), 3843–9.
<https://doi.org/10.1128/jb.174.12.3843-3849.1992>
- Lu, Z., Xie, P., & Qin, Z. (2010). Promotion of markerless deletion of the actinorhodin biosynthetic gene cluster in Streptomyces coelicolor. *Acta Biochimica et Biophysica Sinica*, 42(10), 717–721. <https://doi.org/10.1093/abbs/gmq080>
- Maciag, M., Kochanowska, M., Łyzeń, R., Wegrzyn, G., & Szalewska-Pałasz, A. (2010). ppGpp inhibits the activity of Escherichia coli DnaG primase. *Plasmid*, 63(1), 61–67.
<https://doi.org/10.1016/j.plasmid.2009.11.002>

- MacNeil, D. J., Gewain, K. M., Ruby, C. L., Dezeny, G., Gibbons, P. H., & MacNeil, T. (1992). Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene*, 111(1), 61–68. [https://doi.org/10.1016/0378-1119\(92\)90603-M](https://doi.org/10.1016/0378-1119(92)90603-M)
- Manganelli, R., Dubnau, E., Tyagi, S., Kramer, F. R., & Smith, I. (1999). Differential expression of 10 sigma factor genes in *Mycobacterium tuberculosis*. *Molecular Microbiology*, 31(2), 715–724. <https://doi.org/10.1046/j.1365-2958.1999.01212.x>
- Manganelli, R. (2014). Sigma Factors: Key Molecules in *Mycobacterium tuberculosis* Physiology and Virulence. *Microbiology Spectrum*, 2(1). Retrieved from <http://www.asmscience.org/content/journal/microbiolspec/10.1128/microbiolspec.MG M2-0007-2013>
- Manganelli, R., Provvedi, R., Rodrigue, S., Beaucher, J., Gaudreau, L., Smith, I., & Provvedi, R. (2004). Sigma factors and global gene regulation in *Mycobacterium tuberculosis*. *Journal of Bacteriology*, 186(4), 895–902. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=344228&tool=pmcentrez&rendertype=abstract>
- Manteca, A., & Sanchez, J. (2009). *Streptomyces* development in colonies and soils. *Applied and Environmental Microbiology*, 75(9), 2920–2924. <https://doi.org/10.1128/AEM.02288-08>
- Marr, M. T., & Roberts, J. W. (1997). Promoter recognition as measured by binding of polymerase to nontemplate strand oligonucleotide. *Science (New York, N.Y.)*, 276(5316), 1258–1260. <https://doi.org/10.1126/science.276.5316.1258>
- Mauri, M., & Klumpp, S. (2014). A Model for Sigma Factor Competition in Bacterial Cells. *PLoS Computational Biology*, 10(10). <https://doi.org/10.1371/journal.pcbi.1003845>
- McCormick, J. R., & Flärdh, K. (2012). Signals and regulators that govern *Streptomyces* development. *FEMS Microbiology Reviews*. <https://doi.org/10.1111/j.1574-6976.2011.00317.x>
- McCormick, J. R., Su, E. P., Driks, A., & Losick, R. (1994). Growth and viability of *Streptomyces coelicolor* mutant for the cell division gene *ftsZ*. *Molecular Microbiology*, 14(2), 243–254. <https://doi.org/10.1111/j.1365-2958.1994.tb01285.x>
- McDonald, B. R., & Currie, C. R. (2017). Lateral gene transfer dynamics in the ancient bacterial genus *Streptomyces*. *MBio*, 8(3). <https://doi.org/10.1128/mBio.00644-17>
- McGinness, K. E., Baker, T. A., & Sauer, R. T. (2006). Engineering Controllable Protein Degradation. *Molecular Cell*, 22(5), 701–707. <https://doi.org/10.1016/j.molcel.2006.04.027>
- McMurray, D. N. (1996). *Mycobacteria and Nocardia*. *Medical Microbiology*. <https://doi.org/NBK7812> [bookaccession]

- Meena, L. S., & Rajni. (2010). Survival mechanisms of pathogenic *Mycobacterium tuberculosis* H37Rv. *FEBS Journal*, 277, 2416–2427. <https://doi.org/10.1111/j.1742-4658.2010.07666.x>
- Minakhin, L., Nechaev, S., Campbell, E. A., & Severinov, K. (2001). Recombinant *Thermus aquaticus* RNA polymerase, a new tool for structure-based analysis of transcription. *Journal of Bacteriology*, 183(1), 71–76. <https://doi.org/10.1128/JB.183.1.71-76.2001>
- Minch, K. J., Rustad, T. R., Peterson, E. J. R., Winkler, J., Reiss, D. J., Ma, S., ... Sherman, D. R. (2015). The DNA-binding network of *Mycobacterium tuberculosis*. *Nature Communications*, 6. <https://doi.org/10.1038/ncomms6829>
- Miropolskaya, N., Ignatov, A., Bass, I., Zhilina, E., Pupov, D., & Kulbachinskiy, A. (2012). Distinct functions of regions 1.1 and 1.2 of RNA polymerase σ subunits from *Escherichia coli* and *Thermus aquaticus* in transcription initiation. *Journal of Biological Chemistry*, 287(28), 23779–23789. <https://doi.org/10.1074/jbc.M112.363242>
- Molodtsov, V., Sineva, E., Zhang, L., Huang, X., Cashel, M., Ades, S. E., & Murakami, K. S. (2018). Allosteric Effector ppGpp Potentiates the Inhibition of Transcript Initiation by DksA. *Molecular Cell*, 69(5), 828–839.e5. <https://doi.org/10.1016/j.molcel.2018.01.035>
- Mondal, S., Yakhnin, A. V., Sebastian, A., Albert, I., & Babitzke, P. (2016). NusA-dependent transcription termination prevents misregulation of global gene expression. *Nature Microbiology*, 1(1). <https://doi.org/10.1038/nmicrobiol.2015.7>
- Monteil, V., Kolb, A., Mayer, C., Hoos, S., England, P., & Norel, F. (2010). Crl binds to domain 2 of σ S and confers a competitive advantage on a natural rpoS mutant of *Salmonella enterica* serovar typhi. *Journal of Bacteriology*, 192(24), 6401–6410. <https://doi.org/10.1128/JB.00801-10>
- Moore, S. J., Lai, H. E., Needham, H., Polizzi, K. M., & Freemont, P. S. (2017). *Streptomyces venezuelae* TX-TL – a next generation cell-free synthetic biology tool. *Biotechnology Journal*, 12(4). <https://doi.org/10.1002/biot.201600678>
- Mukamolova, G. V., Turapov, O., Malkin, J., Woltmann, G., & Barer, M. R. (2010). Resuscitation-promoting factors reveal an occult population of tubercle bacilli in sputum. *American Journal of Respiratory and Critical Care Medicine*, 181(2), 174–180. <https://doi.org/10.1164/rccm.200905-0661OC>
- Mukherjee, K., & Chatterji, D. (1997). Studies on the omega subunit of *Escherichia coli* RNA polymerase--its role in the recovery of denatured enzyme activity. *Eur J Biochem*, 247, 884–889. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9288911
- Murakami, K. S., & Darst, S. A. (2003). Bacterial RNA polymerases: The whole story. *Current Opinion in Structural Biology*. [https://doi.org/10.1016/S0959-440X\(02\)00005-2](https://doi.org/10.1016/S0959-440X(02)00005-2)

- Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O., & Darst, S. A. (2002). Structural basis of transcription initiation: An RNA polymerase holoenzyme-DNA complex. *Science*, 296(5571), 1285–1290. <https://doi.org/10.1126/science.1069595>
- Mustaev, a, Zaychikov, E., Severinov, K., Kashlev, M., Polyakov, a, Nikiforov, V., & Goldfarb, a. (1994). Topology of the RNA polymerase active center probed by chimeric rifampicin-nucleotide compounds. *Proceedings of the National Academy of Sciences of the United States of America*, 91(25), 12036–12040. <https://doi.org/10.1073/pnas.91.25.12036>
- Muzzin, O., Campbell, E. A., Xia, L., Severinova, E., Darst, S. A., & Severinov, K. (1998). Disruption of Escherichia coli HepA, an RNA polymerase-associated protein, causes UV sensitivity. *Journal of Biological Chemistry*, 273(24), 15157–15161. <https://doi.org/10.1074/jbc.273.24.15157>
- Narasimhan, P., Wood, J., MacIntyre, C. R., & Mathai, D. (2013). Risk factors for tuberculosis. *Pulmonary Medicine*, 63(1), 37–46. <https://doi.org/10.1155/2013/828939>
- Newell, K. V., Thomas, D. P., Brekasis, D., & Paget, M. S. B. (2006). The RNA polymerase-binding protein RbpA confers basal levels of rifampicin resistance on Streptomyces coelicolor. *Molecular Microbiology*, 60, 687–696. <https://doi.org/10.1111/j.1365-2958.2006.05116.x>
- Newton, G. L., Arnold, K., Price, M. S., Sherrill, C., Delcardayre, S. B., Aharonowitz, Y., ... Davis, C. (1996). Distribution of thiols in microorganisms: Mycothiol is a major thiol in most actinomycetes. *Journal of Bacteriology*, 178(7), 1990–1995. <https://doi.org/10.1021/bi8010253>
- Newton, G. L., & Fahey, R. C. (2008). Regulation of mycothiol metabolism by ??R and the thiol redox sensor anti-sigma factor RsrA. *Molecular Microbiology*. <https://doi.org/10.1111/j.1365-2958.2008.06222.x>
- Newton-Foot, M., & Gey Van Pittius, N. C. (2013). The complex architecture of mycobacterial promoters. *Tuberculosis*. <https://doi.org/10.1016/j.tube.2012.08.003>
- Nguyen, H. T., Wolff, K. A., Cartabuke, R. H., Ogowang, S., & Nguyen, L. (2010). A lipoprotein modulates activity of the MtrAB two-component system to provide intrinsic multidrug resistance, cytokinetic control and cell wall homeostasis in Mycobacterium. *Molecular Microbiology*, 76(2), 348–364. <https://doi.org/10.1111/j.1365-2958.2010.07110.x>
- Nouaille, S., Mondeil, S., Finoux, A.-L., Moulis, C., Girbal, L., & Coccagn-Bousquet, M. (2017). The stability of an mRNA is influenced by its concentration: a potential physical mechanism to regulate gene expression. *Nucleic Acids Research*, 45(20), 11711–11724. <https://doi.org/10.1093/nar/gkx781>
- Ochi, K. (1995). A taxonomic study of the genus Streptomyces by analysis of ribosomal protein AT-L30. *Int J Syst Bacteriol*, 45(3), 507–514. <https://doi.org/10.1099/00207713-45-3-507>

- Paget, M. S., Chamberlin, L., Atrih, A., Foster, S. J., & Buttner, M. J. (1999). Evidence that the extracytoplasmic function sigma factor sigmaE is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). *J Bacteriol*, 181(1), 204–211. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9864331><http://www.ncbi.nlm.nih.gov/pmc/articles/PMC103550/pdf/jb000204.pdf>
- Paget, M. S., Kang, J. G., Roe, J. H., & Buttner, M. J. (1998). sigmaR, an RNA polymerase sigma factor that modulates expression of the thioredoxin system in response to oxidative stress in *Streptomyces coelicolor* A3(2). *The EMBO Journal*, 17(19), 5776–5782. <https://doi.org/10.1093/emboj/17.19.5776>
- Paget, M. S., Leibovitz, E., & Buttner, M. J. (1999). A putative two-component signal transduction system regulates sigmaE, a sigma factor required for normal cell wall integrity in *Streptomyces coelicolor* A3(2). *Molecular Microbiology*, 33(1), 97–107. <https://doi.org/10.1046/j.1365-2958.1999.01452.x>
- Paget, M. S., Molle, V., Cohen, G., Aharonowitz, Y., & Buttner, M. J. (2001). Defining the disulphide stress response in *Streptomyces coelicolor* A3(2): identification of the sigmaR regulon. *Molecular Microbiology*, 42, 1007–1020. <https://doi.org/10.1046/j.1365-2958.2001.02675.x>
- Paget, M. S. B., & Helmann, J. D. (2003). The sigma70 family of sigma factors. *Genome Biology*, 4(1), 203. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=151288&tool=pmcentrez&rendertype=abstract>
- Paget, M. S. (2015). Bacterial sigma factors and anti-sigma factors: Structure, function and distribution. *Biomolecules*. <https://doi.org/10.3390/biom5031245>
- Parshin, A., Shiver, A. L., Lee, J., Ozerova, M., Schneidman-Duhovny, D., Gross, C. A., & Borukhov, S. (2015). DksA regulates RNA polymerase in *Escherichia coli* through a network of interactions in the secondary channel that includes Sequence Insertion 1. *Proceedings of the National Academy of Sciences*, 112(50), E6862–E6871. <https://doi.org/10.1073/pnas.1521365112>
- Paul, B. J., Berkmen, M. B., & Gourse, R. L. (2005). DksA potentiates direct activation of amino acid promoters by ppGpp. *Proceedings of the National Academy of Sciences*, 102(22), 7823–7828. <https://doi.org/10.1073/pnas.0501170102>
- Perederina, A., Svetlov, V., Vassylyeva, M. N., Tahirov, T. H., Yokoyama, S., Artsimovitch, I., & Vassylyev, D. G. (2004). Regulation through the secondary channel - Structural framework for ppGpp-DksA synergism during transcription. *Cell*, 118(3), 297–309. <https://doi.org/10.1016/j.cell.2004.06.030>
- Petrickova, K., & Petricek, M. (2003). Eukaryotic-type protein kinases in *Streptomyces coelicolor*: Variations on a common theme. *Microbiology*. <https://doi.org/10.1099/mic.0.26275-0>

- Pollett, S., Banner, P., O'Sullivan, M. V. N., & Ralph, A. P. (2016). Epidemiology, diagnosis and management of extra-pulmonary tuberculosis in a low-prevalence country: A four year retrospective study in an Australian Tertiary Infectious Diseases Unit. *PLoS ONE*, 11(3). <https://doi.org/10.1371/journal.pone.0149372>
- Public Health England. (2017). Tuberculosis in England 2017 Report. *Public Health England, Version 1.*, 173.
- Pupov, D., Kuzin, I., Bass, I., & Kulbachinskiy, A. (2014). Distinct functions of the RNA polymerase σ subunit region 3.2 in RNA priming and promoter escape. *Nucleic Acids Research*, 42(7), 4494–4504. <https://doi.org/10.1093/nar/gkt1384>
- Qayyum, M. Z., Dey, D., & Sen, R. (2016). Transcription elongation factor NusA is a general antagonist of rho-dependent termination in Escherichia coli. *Journal of Biological Chemistry*, 291(15), 8090–8108. <https://doi.org/10.1074/jbc.M115.701268>
- Rammohan, J., Ruiz Manzano, A., Garner, A. L., Stallings, C. L., & Galburt, E. A. (2015). CarD stabilizes mycobacterial open complexes via a two-tiered kinetic mechanism. *Nucleic Acids Research*, 43(6), 3272–85. <https://doi.org/10.1093/nar/gkv078>
- Rammohan, J., Ruiz Manzano, A., Garner, A. L., Prusa, J., Stallings, C. L., & Galburt, E. A. (2016). Cooperative stabilization of *Mycobacterium tuberculosis* *rrnA* P3 promoter open complexes by RbpA and CarD. *Nucleic Acids Research*, (6), gkw577. <https://doi.org/10.1093/nar/gkw577>
- Ray-Soni, A., Bellecourt, M. J., & Landick, R. (2016). Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annual Review of Biochemistry*, 85(1), 319–347. <https://doi.org/10.1146/annurev-biochem-060815-014844>
- Reichard, P., & Ehrenberg, A. (1983). Ribonucleotide Reductase--a Radical Enzyme. *Source: Science, New Series*, 221(4610), 514–519. <https://doi.org/10.1126/science.6306767>
- Rigali, S., Titgemeyer, F., Barends, S., Mulder, S., Thomae, A. W., Hopwood, D. A., & van Wezel, G. P. (2008). Feast or famine: The global regulator DasR links nutrient stress to antibiotic production by Streptomyces. *EMBO Reports*, 9(7), 670–675. <https://doi.org/10.1038/embo.2008.83>
- Roberts, J. W., Shankar, S., & Filter, J. J. (2008). RNA Polymerase Elongation Factors. *Annual Review of Microbiology*, 62(1), 211–233. <https://doi.org/10.1146/annurev.micro.61.080706.093422>
- Rodrigue, S., Provvedi, R., Jacques, P.-E., Gaudreau, L., & Manganelli, R. (2006). The sigma factors of Mycobacterium tuberculosis. *FEMS Microbiology Reviews*, 30(6), 926–941. <https://doi.org/10.1111/j.1574-6976.2006.00040.x>
- Rodríguez, H., Rico, S., Díaz, M., & Santamaría, R. I. (2013). Two-component systems in Streptomyces: Key regulators of antibiotic complex pathways. *Microbial Cell Factories*. <https://doi.org/10.1186/1475-2859-12-127>

- Ross, W., Sanchez-Vazquez, P., Chen, A. Y., Lee, J. H., Burgos, H. L., & Gourse, R. L. (2016). PpGpp Binding to a Site at the RNAP-DksA Interface Accounts for Its Dramatic Effects on Transcription Initiation during the Stringent Response. *Molecular Cell*, 62(6), 811–823. <https://doi.org/10.1016/j.molcel.2016.04.029>
- Ross, W., Vrentas, C. E., Sanchez-Vazquez, P., Gaal, T., & Gourse, R. L. (2013). The magic spot: A ppGpp binding site on E. coli RNA polymerase responsible for regulation of transcription initiation. *Molecular Cell*, 50(3), 420–429. <https://doi.org/10.1016/j.molcel.2013.03.021>
- Rudd, B. A., & Hopwood, D. A. (1979). Genetics of actinorhodin biosynthesis by *Streptomyces coelicolor* A3(2). *Journal of General Microbiology*, 114(1), 35–43. <https://doi.org/10.1099/00221287-114-1-35>
- Rudd, B. a. M., & Hopwood, D. a. (1980). A pigmented mycelial antibiotic in *Streptomyces coelicolor*: control by a chromosomal gene cluster. *Journal of General Microbiology*, 119(2), 333–340. <https://doi.org/10.1099/00221287-119-2-333>
- Rustad, T. R., Minch, K. J., Ma, S., Winkler, J. K., Hobbs, S., Hickey, M., ... Sherman, D. R. (2014). Mapping and manipulating the *Mycobacterium tuberculosis* transcriptome using a transcription factor overexpression-derived regulatory network. *Genome Biology*, 15(11), 502. <https://doi.org/10.1186/PREACCEPT-1701638048134699>
- Rutledge, P. J., & Challis, G. L. (2015). Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nature Reviews Microbiology*. <https://doi.org/10.1038/nrmicro3496>
- Sachdeva, P., Misra, R., Tyagi, A. K., & Singh, Y. (2010). The sigma factors of *Mycobacterium tuberculosis*: Regulation of the regulators. *FEBS Journal*. <https://doi.org/10.1111/j.1742-4658.2009.07479.x>
- Saecker, R. M., Record, M. T., & Dehaseth, P. L. (2011). Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *Journal of Molecular Biology*, 412(5), 754–71. <https://doi.org/10.1016/j.jmb.2011.01.018>
- Salina, E. G., Waddell, S. J., Hoffmann, N., Rosenkrands, I., Butcher, P. D., & Kaprelyants, A. S. (2014). Potassium availability triggers *Mycobacterium tuberculosis* transition to, and resuscitation from, non-culturable (dormant) states. *Open Biology*, 4(10), 140106–140106. <https://doi.org/10.1098/rsob.140106>
- Salina, E. G., Waddell, S. J., Hoffmann, N., Rosenkrands, I., Butcher, P. D., & Kaprelyants, A. S. (2014). Potassium availability triggers *Mycobacterium tuberculosis* transition to, and resuscitation from, non-culturable (dormant) states. *Open Biology*, 4(10). <https://doi.org/10.1098/rsob.140106>

- Samanta, S., & Martin, C. T. (2013). Insights into the mechanism of initial transcription in *Escherichia coli* RNA polymerase. *The Journal of Biological Chemistry*, 288(44), 31993–2003. <https://doi.org/10.1074/jbc.M113.497669>
- Santos, E. L. de los, & Challis, G. L. (2017). clusterTools: Proximity Searches For Functional Elements To Identify Putative Biosynthetic Gene Clusters. *BioRxiv*, 119214. <https://doi.org/10.1101/119214>
- Santos-Beneit, F., Barriuso-Iglesias, M., Fernández-Martínez, L. T., Martínez-Castro, M., Sola-Landa, A., Rodríguez-García, A., & Martín, J. F. (2011). The RNA polymerase omega factor rpoZ is regulated by phop and has an important role in antibiotic biosynthesis and morphological differentiation in *Streptomyces coelicolor*. *Applied and Environmental Microbiology*, 77(21), 7586–7594. <https://doi.org/10.1128/AEM.00465-11>
- Santos-Beneit, F., Rodríguez-García, A., Sola-Landa, A., & Martín, J. F. (2009). Cross-talk between two global regulators in *Streptomyces*: PhoP and AfsR interact in the control of afsS, pstS and phoRP transcription. *Molecular Microbiology*, 72(1), 53–68. <https://doi.org/10.1111/j.1365-2958.2009.06624.x>
- Sassetti, C. M., Boyd, D. H., & Rubin, E. J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular Microbiology*, 48(1), 77–84. <https://doi.org/10.1046/j.1365-2958.2003.03425.x>
- Sauer, R. T., Bolon, D. N., Burton, B. M., Burton, R. E., Flynn, J. M., Grant, R. A., ... Baker, T. A. (2004). Sculpting the proteome with AAA+ proteases and disassembly machines. *Cell*. <https://doi.org/10.1016/j.cell.2004.09.020>
- Savery, N. J., Lloyd, G. S., Kainz, M., Gaal, T., Ross, W., Ebright, R. H., ... Busby, S. J. W. (1998). Transcription activation at class II CRP-dependent promoters: Identification of determinants in the C-terminal domain of the RNA polymerase α subunit. *EMBO Journal*, 17(12), 3439–3447. <https://doi.org/10.1093/emboj/17.12.3439>
- Schmitz, A., & Galas, D. J. (1979). The interaction of RNA polymerase and lac repressor with the lac control region. *Nucleic Acids Research*, 6(1), 111–137. <https://doi.org/10.1093/nar/6.1.111>
- Schrempf, H. (2001). Recognition and degradation of chitin by streptomycetes. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, 79(3–4), 285–289. <https://doi.org/10.1023/A:1012058205158>
- Schwartz, E. C., Shekhtman, A., Dutta, K., Pratt, M. R., Cowburn, D., Darst, S., & Muir, T. W. (2008). A full-length group 1 bacterial sigma factor adopts a compact structure incompatible with DNA binding. *Chemistry & Biology*, 15(10), 1091–103. <https://doi.org/10.1016/j.chembiol.2008.09.008>
- Scorpio, A., & Zhang, Y. (1996). Mutations in pncA, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculous drug

- pyrazinamide in tubercle bacillus. *Nature Medicine*, 2(6), 662–667.
<https://doi.org/10.1038/nm0696-662>
- Selby, C. P. (2017). Mfd Protein and Transcription–Repair Coupling in *Escherichia coli*. *Photochemistry and Photobiology*. <https://doi.org/10.1111/php.12675>
- Shaw, G., Gan, J., Zhou, Y. N., Zhi, H., Subburaman, P., Zhang, R., ... Ji, X. (2008). Structure of RapA, a Swi2/Snf2 Protein that Recycles RNA Polymerase During Transcription. *Structure*, 16(9), 1417–1427. <https://doi.org/10.1016/j.str.2008.06.012>
- Sheeler, N. L., MacMillan, S. V., & Nodwell, J. R. (2005). Biochemical activities of the absA two-component system of *Streptomyces coelicolor*. *Journal of Bacteriology*, 187(2), 687–696. <https://doi.org/10.1128/JB.187.2.687-696.2005>
- Sherman, D. R., Voskuil, M., Schnappinger, D., Liao, R., Harrell, M. I., & Schoolnik, G. K. (2001). Regulation of the *Mycobacterium tuberculosis* hypoxic response gene encoding alpha - crystallin. *Proceedings of the National Academy of Sciences of the United States of America*, 98(13), 7534–9. <https://doi.org/10.1073/pnas.121172498>
- Sigle, S., Ladwig, N., Wohlleben, W., & Muth, G. (2015). Synthesis of the spore envelope in the developmental life cycle of *Streptomyces coelicolor*. *International Journal of Medical Microbiology*. <https://doi.org/10.1016/j.ijmm.2014.12.014>
- Slepecky, R. A., & Hemphill, H. E. (2006). The Genus *Bacillus*—Nonmedical. In *The Prokaryotes* (pp. 530–562). https://doi.org/10.1007/0-387-30744-3_16
- Smith, I. (2003). *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. *Clinical Microbiology Reviews*. <https://doi.org/10.1128/CMR.16.3.463-496.2003>
- Srivastava, D. B., Leon, K., Osmundson, J., Garner, A. L., Weiss, L. A., Westblade, L. F., ... Campbell, E. A. (2013). Structure and function of CarD, an essential mycobacterial transcription factor. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 12619–12624. <https://doi.org/10.1073/pnas.1308270110>
- Stallings, C. L., Stephanou, N. C., Chu, L., Hochschild, A., Nickels, B. E., & Glickman, M. S. (2009). CarD is an essential regulator of rRNA transcription required for *Mycobacterium tuberculosis* persistence. *Cell*, 138(1), 146–59. <https://doi.org/10.1016/j.cell.2009.04.041>
- Sudalaiyadum Perumal, A., Vishwakarma, R. K., Hu, Y., Morichaud, Z., & Brodolin, K. (2018). RbpA relaxes promoter selectivity of *M. tuberculosis* RNA polymerase. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gky714>
- Sukhodolets, M. V., Cabrera, J. E., Zhi, H., & Ding Jun Jin. (2001). RapA, a bacterial homolog of SWI2/SNF2, stimulates RNA polymerase recycling in transcription. *Genes and Development*, 15(24), 3330–3341. <https://doi.org/10.1101/gad.936701>

- Sun, J., Hesketh, A., & Bibb, M. (2001). Functional analysis of *relA* and *rshA*, two *relA*/*spoT* homologues of *Streptomyces coelicolor* A3(2). *Journal of Bacteriology*, 183(11), 3488–3498. <https://doi.org/10.1128/JB.183.11.3488-3498.2001>
- Sutton, S. (2010). The Most Probable Number Method and its uses in enumeration, qualification, and validation. *Journal of Validation Technology*, 16(3), 35–38. <https://doi.org/10.1097/RLI.0b013e318234e75b.Compartmental>
- Tabib-Salazar, A., Liu, B., Doughty, P., Lewis, R. A., Ghosh, S., Parsy, M. L., ... Paget, M. S. (2013). The actinobacterial transcription factor RbpA binds to the principal sigma subunit of RNA polymerase. *Nucleic Acids Research*, 41, 5679–5691. <https://doi.org/10.1093/nar/gkt277>
- Takano, E., Gramajo, H. C., Strauch, E., Andres, N., White, J., & Bibb, M. J. (1992). Transcriptional regulation of the *redD* transcriptional activator gene accounts for growth-phase-dependent production of the antibiotic undecylprodigiosin in *Streptomyces coelicolor* A3(2). *Molecular Microbiology*, 6(19), 2797–2804. <https://doi.org/10.1111/j.1365-2958.1992.tb01459.x>
- Takano, H., Obitsu, S., Beppu, T., & Ueda, K. (2005). Light-induced carotenogenesis in *Streptomyces coelicolor* A3(2): Identification of an extracytoplasmic function sigma factor that directs photodependent transcription of the carotenoid biosynthesis gene cluster. *Journal of Bacteriology*, 187(5), 1825–1832. <https://doi.org/10.1128/JB.187.5.1825-1832.2005>
- Tetsuo, S., Kan, T., & Hideo, T. (1991). Sequence of *hrdB*, an essential gene encoding sigma-like transcription factor of *Streptomyces coelicolor* A3(2): homology to principal sigma factors. *Gene*, 107(1), 145–148. [https://doi.org/10.1016/0378-1119\(91\)90308-X](https://doi.org/10.1016/0378-1119(91)90308-X)
- Thakur, K. G., Joshi, A. M., & Gopal, B. (2007). Structural and biophysical studies on two promoter recognition domains of the extra-cytoplasmic function σ factor σ_C from *Mycobacterium tuberculosis*. *Journal of Biological Chemistry*, 282(7), 4711–4718. <https://doi.org/10.1074/jbc.M606283200>
- Thakur, S., Normand, P., Daubin, V., Tisa, L. S., & Sen, A. (2013). Contrasted evolutionary constraints on secreted and non-secreted proteomes of selected Actinobacteria. *BMC Genomics*, 14, 474. <https://doi.org/10.1186/1471-2164-14-474>
- The World Health Organization. (2010). Treatment of Tuberculosis: Guidelines. <https://doi.org/10.1164/rccm.201012-1949OC>
- Tiffert, Y., Supra, P., Wurm, R., Wohlleben, W., Wagner, R., & Reuther, J. (2008). The *Streptomyces coelicolor* GlnR regulon: Identification of new GlnR targets and evidence for a central role of GlnR in nitrogen metabolism in actinomycetes. *Molecular Microbiology*, 67(4), 861–880. <https://doi.org/10.1111/j.1365-2958.2007.06092.x>
- Trapnell, C., Williams, B. a, Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., ... Pachter, L. (2011). Transcript assembly and abundance estimation from RNA-Seq reveals thousands

- of new transcripts and switching among isoforms. *Nature Biotechnology*, 28(5), 511–515. <https://doi.org/10.1038/nbt.1621>. Transcript
- Treviño-Quintanilla, L. G., Freyre-González, J. A., & Martínez-Flores, I. (2013). Anti-Sigma Factors in *E. coli*: Common Regulatory Mechanisms Controlling Sigma Factors Availability. *Current Genomics*, 14(6), 378–87. <https://doi.org/10.2174/1389202911314060007>
- Tudó, G., Laing, K., Mitchison, D. A., Butcher, P. D., & Waddell, S. J. (2010). Examining the basis of isoniazid tolerance in nonreplicating *Mycobacterium tuberculosis* using transcriptional profiling. *Future Medicinal Chemistry*, 2(8), 1371–1383. <https://doi.org/10.4155/fmc.10.219>
- Tundup, S., Akhter, Y., Thiagarajan, D., & Hasnain, S. E. (2006). Clusters of PE and PPE genes of *Mycobacterium tuberculosis* are organized in operons: Evidence that PE Rv2431c is co-transcribed with PPE Rv2430c and their gene products interact with each other. *FEBS Letters*, 580(5), 1285–1293. <https://doi.org/10.1016/j.febslet.2006.01.042>
- Turkarslan, S., Peterson, E. J. R., Rustad, T. R., Minch, K. J., Reiss, D. J., Morrison, R., ... Baliga, N. S. (2015). A comprehensive map of genome-wide gene regulation in *Mycobacterium tuberculosis*. *Scientific Data*, 2. <https://doi.org/10.1038/sdata.2015.10>
- Uguru, G. C., Stephens, K. E., Stead, J. A., Towle, J. E., Baumberg, S., & McDowall, K. J. (2005). Transcriptional activation of the pathway-specific regulator of the actinorhodin biosynthetic genes in *Streptomyces coelicolor*. *Molecular Microbiology*, 58(1), 131–150. <https://doi.org/10.1111/j.1365-2958.2005.04817.x>
- Van Wezel, G. P., & Bibb, M. J. (1996). A novel plasmid vector that uses the glucose kinase gene (*glkA*) for the positive selection of stable gene disruptants in *Streptomyces*. *Gene*, 182(1–2), 229–230. [https://doi.org/10.1016/S0378-1119\(96\)00563-X](https://doi.org/10.1016/S0378-1119(96)00563-X)
- van Wezel, G. P., & McDowall, K. J. (2011). The regulation of the secondary metabolism of *Streptomyces*: new links and experimental advances. *Natural Product Reports*, 28(7), 1311. <https://doi.org/10.1039/c1np00003a>
- Vassilyev, D. G., Vassilyeva, M. N., Perederina, A., Tahirov, T. H., & Artsimovitch, I. (2007). Structural basis for transcription elongation by bacterial RNA polymerase. *Nature*, 448(7150), 157–162. <https://doi.org/10.1038/nature05932>
- Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G. F., Chater, K. F., & van Sinderen, D. (2007). Genomics of Actinobacteria: Tracing the Evolutionary History of an Ancient Phylum. *Microbiology and Molecular Biology Reviews*, 71(3), 495–548. <https://doi.org/10.1128/MMBR.00005-07>
- Verbon, A., Hartskeerl, R. A., Schuitema, A., Kolk, A. H. J., Young, D. B., & Lathigra, R. (1992). The 14,000-molecular-weight antigen of *Mycobacterium tuberculosis* is related to the alpha-crystallin family of low-molecular-weight heat shock proteins. *Journal of Bacteriology*, 174(4), 1352–1359.

- Vilchèze, C., Hartman, T., Weinrick, B., & Jacobs, W. R. (2013). Mycobacterium tuberculosis is extraordinarily sensitive to killing by a vitamin C-induced Fenton reaction. *Nature Communications*, 4, 1881. <https://doi.org/10.1038/ncomms2898>
- Vishwakarma, R. K., Cao, A., Morichaud, Z., Perumal, A. S., Margeat, E., & Brodolin, K. (2018). Single-molecule analysis reveals the mechanism of transcription activation in M. tuberculosis. *Science Advances*, (May), 1–9. <https://doi.org/10.1126/sciadv.aao5498>
- Wang, Z., Cumming, B. M., Mao, C., Zhu, Y., Lu, P., Steyn, A. J. C., ... Hu, Y. (2018). RbpA and σ B association regulates polyphosphate levels to modulate mycobacterial isoniazid-tolerance. *Molecular Microbiology*, 0(0). <https://doi.org/10.1111/mmi.13952>
- Washburn, R. S., & Gottesman, M. E. (2015). Regulation of transcription elongation and termination. *Biomolecules*. <https://doi.org/10.3390/biom5021063>
- Wawrzynow, a, Wojtkowiak, D., Marszalek, J., Banecki, B., Jonsen, M., Graves, B., ... Zylicz, M. (1995). The ClpX heat-shock protein of Escherichia coli, the ATP-dependent substrate specificity component of the ClpP-ClpX protease, is a novel molecular chaperone. *The EMBO Journal*, 14(9), 1867–1877.
- Weber, H., Polen, T., Heuveling, J., Wendisch, V. F., & Hengge, R. (2005). Genome-wide analysis of the general stress response network in *Escherichia coli*: σ S-Dependent genes, promoters, and sigma factor selectivity. *Journal Of Bacteriology*, 187(5), 1591–1603. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1063999&tool=pmcentrez&rendertype=abstract>
- Weiss, L. A., Harrison, P. G., Nickels, B. E., Glickman, M. S., Campbell, E. A., Darst, S. A., & Stallings, C. L. (2012). Interaction of CarD with RNA polymerase mediates Mycobacterium tuberculosis viability, rifampin resistance, and pathogenesis. *Journal of Bacteriology*, 194(20), 5621–5631. <https://doi.org/10.1128/JB.00879-12>
- Westpheling, J., Raney, M., & Losick, R. (1985). RNA polymerase heterogeneity in Streptomyces coelicolor. *Nature*, 313(5997), 22–27. <https://doi.org/10.1038/313022a0>
- WHO. (2017). *WHO Global tuberculosis report 2017*. World Health Organization Press. <https://doi.org/ISBN 978 92 4 156539 4>
- Wiatkiewicz-Połatowska, M. A., Bucca, G., Laing, E., Gubbens, J., Titgemeyer, F., Smith, C. P., ... Van Wezel, G. P. (2015). Genome-wide analysis of in vivo binding of the master regulator DasR in streptomyces coelicolor identifies novel non-canonical targets. *PLoS ONE*, 10(4). <https://doi.org/10.1371/journal.pone.0122479>
- Wigneshweraraj, S., Bose, D., Burrows, P. C., Joly, N., Schumacher, J., Rappas, M., ... Buck, M. (2008). Modus operandi of the bacterial RNA polymerase containing the sigma54 promoter-specificity factor. *Molecular Microbiology*, 68(3), 538–46. <https://doi.org/10.1111/j.1365-2958.2008.06181.x>

- Winder, F. G., & Collins, P. B. (1970). Inhibition by Isoniazid of Synthesis of Mycolic Acids in *Mycobacterium tuberculosis*. *Journal of General Microbiology*, 63(1), 41–48. <https://doi.org/10.1099/00221287-63-1-41>
- Wolanski, M., Wali, R., Tilley, E., Jakimowicz, D., Zakrzewska-Czerwińska, J., & Herron, P. (2011). Replisome trafficking in growing vegetative hyphae of *Streptomyces coelicolor* A3(2). *Journal of Bacteriology*, 193(5), 1273–1275. <https://doi.org/10.1128/JB.01326-10>
- Wolfgang Ludwig and Hans-Peter Klenk. (n.d.). Overview: A Phylogenetic Backbone and Taxonomic Framework for Prokaryotic Systematics. Retrieved June 1, 2015, from http://nebc.nerc.ac.uk/downloads/courses/ARB/prokaryote_phylogeny.pdf
- Wray, L. V., & Fisher, S. H. (1993). The *Streptomyces coelicolor* *glnR* gene encodes a protein similar to other bacterial response regulators. *Gene*, 130(1), 145–150. [https://doi.org/10.1016/0378-1119\(93\)90359-B](https://doi.org/10.1016/0378-1119(93)90359-B)
- WRIGHT, L. F., & HOPWOOD, D. A. (1976). Actinorhodin is a Chromosomally-determined Antibiotic in *Streptomyces coelicolor* A3(2). *Journal of General Microbiology*, 96(2), 289–297. <https://doi.org/10.1099/00221287-96-2-289>
- Wu, S., Howard, S. T., Lakey, D. L., Kipnis, A., Samten, B., Safi, H., ... Barnes, P. F. (2004). The principal sigma factor *sigA* mediates enhanced growth of *Mycobacterium tuberculosis* in vivo. *Molecular Microbiology*, 51(6), 1551–1562. <https://doi.org/10.1111/j.1365-2958.2003.03922.x>
- Wu, X., Haakonsen, D. L., Sanderlin, A. G., Liu, Y. J., Shen, L., Zhuang, N., ... Zhang, Y. (2018). Structural insights into the unique mechanism of transcription activation by *Caulobacter crescentus* GcrA. *Nucleic Acids Research*, 46(6), 3245–3256. <https://doi.org/10.1093/nar/gky161>
- Yakhnin, A. V., & Babitzke, P. (2010). Mechanism of NusG-stimulated pausing, hairpin-dependent pause site selection and intrinsic termination at overlapping pause and termination sites in the *Bacillus subtilis* *trp* leader. *Molecular Microbiology*, 76(3), 690–705. <https://doi.org/10.1111/j.1365-2958.2010.07126.x>
- Yakhnin, A. V., Murakami, K. S., & Babitzke, P. (2016). NusG is a sequence-specific RNA polymerase pause factor that binds to the non-template DNA within the paused transcription bubble. *Journal of Biological Chemistry*, 291(10), 5299–5308. <https://doi.org/10.1074/jbc.M115.704189>
- Yao, L. L., & Ye, B. C. (2016). Reciprocal regulation of *GlnR* and *PhoP* in response to nitrogen and phosphate limitations in *Saccharopolyspora erythraea*. *Applied and Environmental Microbiology*, 82(1), 409–420. <https://doi.org/10.1128/AEM.02960-15>
- Young, D. B., Comas, I., & de Carvalho, L. P. S. (2015). Phylogenetic analysis of vitamin B12-related metabolism in *Mycobacterium tuberculosis*. *Frontiers in Molecular Biosciences*, 2. <https://doi.org/10.3389/fmolb.2015.00006>

- Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K., & Darst, S. A. (1999). Crystal structure of thermus aquaticus core RNA polymerase at 3.3 Å resolution. *Cell*, 98(6), 811–824. [https://doi.org/10.1016/S0092-8674\(00\)81515-9](https://doi.org/10.1016/S0092-8674(00)81515-9)
- Zhang, Y., Mooney, R. A., Grass, J. A., Sivaramakrishnan, P., Herman, C., Landick, R., & Wang, J. D. (2014). DksA Guards Elongating RNA Polymerase against Ribosome-Stalling-Induced Arrest. *Molecular Cell*, 53(5), 766–778. <https://doi.org/10.1016/j.molcel.2014.02.005>
- Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M. X., Arnold, E., & Ebright, R. H. (2012). Structural basis of transcription initiation. *Science (New York, N.Y.)*, 338(6110), 1076–80. <https://doi.org/10.1126/science.1227786>
- Zhu, M., Dai, X., Guo, W., Ge, Z., Yang, M., Wang, H., & Wang, Y. P. (2017). Manipulating the bacterial cell cycle and cell size by titrating the expression of ribonucleotide reductase. *MBio*, 8(6). <https://doi.org/10.1128/mBio.01741-17>
- Zuo, Y., & Steitz, T. A. (2015). Crystal Structures of the E. coli Transcription Initiation Complexes with a Complete Bubble. *Molecular Cell*, 58(3), 534–540. <https://doi.org/10.1016/j.molcel.2015.03.010>
- WHO | Tuberculosis (TB). (n.d.). Retrieved from <http://www.who.int/gho/tb/en/>